DOCUMENT RESUME

ED 221 561 TM 820 583

TITLF	The Educational Testing Act of 1981. Joint Hearings before the Subcommittee on Elementary, Secondary and Vocational Education and the Subcommittee on Postsecondary Education of the Committee on Education and Labor. House of Representatives, Ninety-Seventh Congress, First Session (July 21, 22, November 4, 5, 1981) on H.R. 1662: To Require Certain Information Be Provided to Individuals Who Take Standardized Educational Admissions Tests, and for Other Purposes.
INSTITUTION	Congress of the U.S., Washington, D.C. House Committee on Education and Labor.
PUB DATE	82
NOTE	1,086p.; Many pages are marginally legible due to small print.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402.
EDRS PRICE	MF08 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Admission Criteria; *Aptitude Tests; *College Entrance Examinations; *Disclosure; *Hearings; High Schools; *Information Dissemination; Legislation; Postsecondary Education; Privacy; Public Policy; Student Rights; Test Use
IDENTIFIERS	Congress 97th; *Testing Legislation; *Truth in Testing Legislation

ABSTRACT

The hearings on a House bill to require that certain test information be disclosed to individuals who take standardized educational admissions tests are presented. The text of H.R. 1662 is given, which specifies the information to be provided to individuals and postsecondary institutions; regulates the reports and statistical data to be required; insures the privacy of test scores; and states guidelines for testing costs. The verbal record of the witnesses before the Subcommittee is given with any prepared statements, letters and supplemental materials. Persons from the testing industry, public school systems, postsecondary institutions, and educational and professional organizations discuss the state of educational testing and present policies and activities in test disclosure and information use. Samples of actual tests, reports and publications on the policies of relevant organizations and examples of studies regarding educational testing are provided. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED221561

TM

THE EDUCATIONAL TESTING ACT OF 1981

JOINT HEARINGS

BEFORE THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

AND THE

SUBCOMMITTEE ON POSTSECONDARY EDUCATION

OF THE

COMMITTEE ON EDUCATION AND LABOR
HOUSE OF REPRESENTATIVES

NINETY-SEVENTH CONGRESS

FIRST SESSION

ON

H.R. 1662

TO REQUIRE CERTAIN INFORMATION BE PROVIDED TO INDIVIDUALS WHO TAKE STANDARDIZED EDUCATIONAL ADMISSIONS TESTS, AND FOR OTHER PURPOSES

HEARINGS HELD IN WASHINGTON, D.C., ON JULY 21, 22; NOVEMBER 4 AND 5, 1981

Printed for the use of the Committee on Education and Labor

TM 820 583

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

THE EDUCATIONAL TESTING ACT OF 1981

JOINT HEARINGS

BEFORE THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

AND THE

SUBCOMMITTEE ON POSTSECONDARY EDUCATION

OF THE

COMMITTEE ON EDUCATION AND LABOR
HOUSE OF REPRESENTATIVES

NINETY-SEVENTH CONGRESS

FIRST SESSION

ON

H.R. 1662

TO REQUIRE CERTAIN INFORMATION BE PROVIDED TO INDIVIDUALS WHO TAKE STANDARDIZED EDUCATIONAL ADMISSIONS TESTS, AND FOR OTHER PURPOSES

HEARINGS HELD IN WASHINGTON, D.C., ON JULY 21, 22;
NOVEMBER 4 AND 5, 1981

Printed for the use of the Committee on Education and Labor

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON : 1982

COMMITTEE ON EDUCATION AND LABOR

CARL D. PERKINS, Kentucky, *Chairman*

AUGUSTUS F. HAWKINS, California
WILLIAM D. FORD, Michigan
PHILLIP BURTON, California
JOSEPH M. GAYDOS, Pennsylvania
WILLIAM (BILL) CLAY, Missouri
MARIO BIAGGI, New York
IKE ANDREWS, North Carolina
PAUL SIMON, Illinois
GEORGE MILLER, California
AUSTIN J. MURPHY, Pennsylvania
TED WEISS, New York
BALTASAR CORRADA, Puerto Rico
DALE E. KILDEE, Michigan
PETER A. PEYSER, New York
PAT WILLIAMS, Montana
WILLIAM R. RATCHFORD, Connecticut
RAY KOGOVSEK, Colorado
HAROLD WASHINGTON, Illinois
DENNIS E. ECKART, Ohio

JOHN M. ASHBROOK, Ohio
JOHN N. ERLENBORN, Illinois
JAMES M. JEFFORDS, Vermont
WILLIAM F. GOODLING, Pennsylvania
E. THOMAS COLEMAN, Missouri
KEN KRAMER, Colorado
ARLEN ERDAHL, Minnesota
THOMAS E. PETRI, Wisconsin
MILLICENT FENWICK, New Jersey
MARGE ROUKEMA, New Jersey
EUGENE JOHNSTON, North Carolina
LAWRENCE J. DeNARDIS, Connecticut
LARRY E. CRAIG, Idaho
WENDELL BAILEY, Missouri

Subcommittee on Elementary, Secondary, and Vocational Education

CARL D. PERKINS, Kentucky, *Chairman*

WILLIAM D. FORD, Michigan
IKE ANDREWS, North Carolina
GEORGE MILLER, California
BALTASAR CORRADA, Puerto Rico
DALE E. KILDEE, Michigan
PAT WILLIAMS, Montana
AUGUSTUS F. HAWKINS, California
MARIO BIAGGI, New York
WILLIAM R. RATCHFORD, Connecticut
HAROLD WASHINGTON, Illinois
—— ——

WILLIAM F. GOODLING, Pennsylvania
JAMES M. JEFFORDS, Vermont
E. THOMAS COLEMAN, Missouri
ARLEN ERDAHL, Minnesota
THOMAS E. PETRI, Wisconsin
(Ex Officio)
MARGE ROUKEMA, New Jersey
LAWRENCE J. DeNARDIS, Connecticut
LARRY E. CRAIG, Idaho

Subcommittee on Postsecondary Education

PAUL SIMON, Illinois, *Chairman*

WILLIAM D. FORD, Michigan
PETER A. PEYSER, New York
JOSEPH M. GAYDOS, Pennsylvania
TED WEISS, New York
IKE ANDREWS, North Carolina
DENNIS E. ECKART, Ohio
CARL D. PERKINS, Kentucky
(Ex Officio)

E. THOMAS COLEMAN, Missouri
JOHN N. ERLENBORN, Illinois
ARLEN ERDAHL, Minnesota
(Ex Officio)
LAWRENCE J. DeNARDIS, Connecticut
WENDELL BAILEY, Missouri

CONTENTS

Page

Text of H.R. 1662 ... 2

Hearings held in Washington, D.C. on:

- July 21, 1981 ... 1
- July 22, 1981 ... 171
- November 4, 1981 ... 391
- November 5, 1981 ... 807

Statement of—

- Anrig, Gregory, president, Educational Testing Services ... 393
- Austen, Mary Ann, counsel to State Senator Kenneth P. LaValle of New York ... 760
- Bennett, Carolyn, student, Johns Hopkins University ... 172
- Casteen, John T., III, dean of admissions, University of Virginia ... 197
- Cooper, John, president, Association of Medical Colleges ... 769
- Galligan, Michael, graduate, Clarkstown South High School, New City, N.Y ... 173
- Goslin, David A., executive director, Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, D.C ... 832
- Hanford, George, president, the College Board ... 52
- Johnson, Sylvia, professor of research methodology, School of Education, Department of Psychoeducational Studies, Howard University ... 36
- Kaplan, Stanley H., founder, Stanley H. Kaplan Educational Center ... 877
- Mehrens, William, professor of measurement and statistics, School of Education, Department of Educational Psychology, Michigan State University ... 18
- Pines, Lois, former Boston Regional Director, Federal Trade Commission ... 808
- Priddy, Michael D., director of research, planning, and evaluation, Guilford County School System, Greensboro, N.C., accompanied by Marian Epstein, expert in test development ... 179
- Rever, Dr. Philip R., director, American College Testing Service, Washington, D.C ... 586
- Robinson, Sharon, director, Instruction and Professional Development, National Education Association ... 224
- Schafer, William, professor, Educational Measurement and Statistics, College of Education, Department of Measurement and Statistics, University of Maryland ... 194
- Shields, Patrick, director, East Harlem College and Career Counseling Program, New York, N.Y ... 182
- Sjogren, Cliff, director of admissions, University of Michigan ... 216
- Wigdor, Alexandra, study director, Committee on Ability Testing, National Academy of Sciences, Washington, D.C ... 838
- Zimmer, Bruce, executive director, Law School Admission Council ... 59

Prepared statements, letters, supplemental materials, etc.—

- American Dental Association, prepared statement of ... 800

Anrig, Gregory R., president, Educational Testing Service, Princeton, N.J.:

- Appendix A.—Board of Trustees Policy on Public Issues ... 408
- Appendix B.—Public Interest Principles for the Design and Use of Admissions Testing Programs ... 411
- Appendix C.—ETS Standards for Quality and Fairness ... 415
- Appendix D.—Summary Report on Test Disclosure Requests, Test Item Challenges, Hand-Scoring Requests, and Test Fees ... 453
- Appendix E.—Members of ETS Board of Trustees, 1980–81 ... 455
- "ETS in Fact, 1981," pamphlet entitled ... 484

Prepared statements, letters, supplemental materials, etc.—Continued
Anrig, Gregory R., president, Educational Testing Service, Princeton, N.J.—Continued
"Issues of Effectiveness and Equity in the Coaching Controversy: Implications for Educational and Testing Practice," Research Report ..... Page 509
Letter to Chairman Perkins, dated December 10, 1981, enclosing requested additional information 461
Prepared statement of 399
"Preparing for Tests," booklet entitled 583
"Score Validity Procedures," pamphlet entitled 474
"Side-by-Side Comparison of H.R. 1662 Provisions and Current Practices of Guidance and Admissions Tests Administered by ETS," pamphlet entitled 496
Association of American Publishers, Inc., prepared statement of 900
Austen, Mary Ann M., counsel to New York State Senator Kenneth P. LaValle, prepared statement of 763
Baer, Joe, director of guidance, South Milwaukee Senior High School, South Milwaukee, Wis., letter to Congressman Clement J. Zablocki, enclosing a letter 1079
Casteen, John T., III., dean of admissions, University of Virginia:
"On the Competencies Students Ought to Have," article entitled 215
Prepared statement of 201
Proposal for the National Association of College Admissions Counselors Forum on College Admissions Tests 205
Cooper, John, president, Association of American Medical Colleges, prepared statement of 776
Dickinson, Earl O., D.V.M., president, the Association of American Veterinary Medical Colleges:
Letter to Chairman Perkins, dated July 24, 1981 385
Prepared statement of 386
Epstein, Marion G., senior adviser, Educational Testing Service, Princeton, N.J.:
Attachment A, "Taking the SAT, a Guide to the Scholastic Aptitude Test and the Test of Standard Written English," article entitled 304
Attachment B, Standardized Testing Has Become Education's Latest Scapegoat," article from the American School Board Journal 352
Attachment C, "Developing a Test," pamphlet entitled 358
Attachment D, "Questioning the Questions, the Process for Challenging a Question on an ETS Test," 360
Letter to Chairman Simon, enclosing additional information, dated August 6, 1981 299
Goslin, David A., executive director, Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, D.C., prepared statement of 834
Hanford, George H., president, the College Board:
Prepared statement of 55
Letter to Chairman Perkins, enclosing supplemental statement, dated July 31, 1981 110
Johnson, Sylvia, professor of research methodology, School of Education, Department of Psychoeducational Studies, Howard University, prepared statement of 39
Kaplan, Stanley H., founder, Stanley H. Kaplan Educational Center, prepared statement of 869
Mehrens, William, professor of measurement and statistics, School of Education, Department of Educational Psychology, Michigan State University:
"ETS Versus Nairn/Nader: Who Reigns?," article from Measurement and Evaluation in Guidance, vol. 14, No. 2, July 1981 26
Letter to Congressman Simon, enclosing responses to request for additional information, dated July 27, 1981 108
"NCME Statement on Educational Admissions Testing," article entitled 23
Messick, Dr. Samuel, vice president and distinguished research scientist, Educational Testing Service, Princeton, N.J.:
Letter from Congressman William F. Goodling, requesting information, dated November 16, 1981 918

Prepared statements, letters, supplemental materials, etc.—Continued

- Messick, Dr. Samuel, vice president and distinguished research scientist, Educational Testing Service, Princeton, N.J.—Continued
  - Letter to Congressman William F. Goodling, answering additional questions, dated November 30, 1981 ..... Page 919
  - "The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research From the Fifties to the FTC," report entitled 925
  - "Time and Method in Coaching for the SAT," excerpt from Psychological Bulletin, 1981 1051
- Miller, James C., III, chairman, Federal Trade Commission, Washington, D.C., letter to Chairman Simon, in response to questions concerning FTC reports, dated December 7, 1981 822
- Muris, Timothy J., director, Bureau of Consumer Protection, Federal Trade Commission, prepared statement of 914
- Peltason, J. W., Office of the President, American Council on Education, Washington, D.C., letter to Chairman Perkins, dated August 20, 1981 379
- Pines, Lois G., former Director, Boston Regional Office, Federal Trade Commission, prepared statement of 812
- Ravitch, Diane, Division of Philosophy, the Social Sciences, and Education, Teachers College, Columbia University:
  - "Forgetting the Questions, the Problem of Educational Reform," article from the American Scholar 367
  - Letter to Chairman Simon, dated July 13, 1981 365
- Rever, Dr. Philip R., the American College Testing Program:
  - ACT Research Reports:
    - No. 6, August 1965 707
    - No. 69, August 1975 723
    - No. 80, October 1979 738
  - Prepared statement of 587
  - The ACT Assessment, sample test booklet:
    - 8120C 599
    - 8121C 635
    - 8121B 671
- Robinson, Sharon, director, Instruction and Professional Development, National Education Association:
  - "Survey of School Programs and Practices 1980," research memo entitled 225
  - Prepared statement of 275
- Rudley, Suanne, policy analyst, Office of Government Relations, the National PTA, Washington, D.C., letter to Buddy Blakey, enclosing prepared statement, dated November 13, 1981 915
- Sammons, James H., M.D., executive vice president, American Medical Association, Chicago, Ill., letter to Chairman Perkins, dated August 4, 1981 383
- Shields, Patrick, M., associate director, East Harlem College, and Career Counseling Program, New York City, prepared statement of 186
- Simon, Hon. Paul, a Representative in Congress from the State of Illinois and Chairman of the Subcommittee on Postsecondary Education:
  - Opening statement 391
  - Opening statement of, November 5, 1981 807
- Sjogren, Dr. Cliff, director of admissions, University of Michigan, prepared statement of 220
- Smith, Kathleen D., RDH, MS, president, American Dental Hygienists' Association, Washington, D.C. letter to Chairman Perkins, dated November 25, 1981 1077
- Solomon, Steve, co-director, the Committee for Fair and Open Testing, New York, N.Y.:
  - Appendix A 136
  - Appendix B 139
  - Appendix C 148
  - Appendix D 167
  - Testimony of 118
- Wigdnor, Alexandra K., on behalf of the Committee on Ability Testing, National Academy of Sciences:
  - Committee on Ability Testing (list of members) 855
  - Prepared statement of 842

Prepared statements, letters, supplemental materials, etc.—Continued

Zimmer, Bruce I., executive director, Law School Admission Council, Washington, D.C.: Page
  Appendix A 71
  Appendix B 77
  Law School Admission Bulletin:
    November 1980 72
    January 1981 76
    July 1981 81
  Prepared statement of 63

8

THE EDUCATIONAL TESTING ACT OF 1981

TUESDAY, JULY 21, 1981

HOUSE OF REPRESENTATIVES, SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION JOINTLY WITH THE SUBCOMMITTEE ON POSTSECONDARY EDUCATION, COMMITTEE ON EDUCATION AND LABOR,

Washington, D.C.

The subcommittees met, pursuant to call, at 9:30 a.m., in room 2175 of the Rayburn House Office Building, Hon. Carl Perkins (chairman of the full committee) presiding.

Members present: Representatives Perkins, Simon, Peyser, Weiss, Erdahl, and Craig.

Staff present: John F. Jennings, counsel; and William Blakey, counsel; Nancy L. Kober, legislative specialist.

Chairman **PERKINS**. The Subcommittee on Elementary, Secondary, and Vocational Education and the Subcommittee on Postsecondary Education are conducting a joint hearing this morning on H.R. 1662.

[Text of H.R. 1662 follows:]

97th CONGRESS
1st Session

H. R. 1662

To require certain information be provided to individuals who take standardized educational admissions tests, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

February 4, 1981

Mr. Weiss (for himself, Mrs. Chisholm, Mr. Miller of California, Mr. Gibbons, Mr. Moffett, Mr. Richmond, Mrs. Schroeder, Mr. Walgren, and Mr. Dixon) introduced the following bill; which was referred to the Committee on Education and Labor

---

A BILL

To require certain information be provided to individuals who take standardized educational admissions tests, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SHORT TITLE

Section 1. This Act may be cited as the "Educational Testing Act of 1981".

FINDINGS AND PURPOSE

SEC. 2. (a) The Congress of the United States finds that—

(1) education is fundamental to the development of individual citizens and the progress of the Nation as a whole;

(2) there is a continuous need to ensure equal access for all Americans to educational opportunities of a high quality;

(3) standardized tests are a major factor in the admission and placement of students in postsecondary education and also play an important role in individuals' professional lives;

(4) there is increasing concern among citizens, educators, and public officials regarding the appropriate uses of standardized tests in the admissions decision of postsecondary education institutions;

(5) the rights of individuals and the public interest can be assured without endangering the proprietary rights of the testing agencies; and

(6) standardized tests are developed and administered without regard to State boundaries and are utilized on a national basis.

(b) It is the purpose of this Act—

(1) to ensure that test subjects and persons who use test results are fully aware of the characteristics, uses, and limitations of standardized tests in postsecondary education admissions;

(2) to make available to the public appropriate information regarding the procedures, development, and administration of standardized tests;

(3) to protect the public interest by promoting more knowledge about appropriate use of standardized test results and by promoting greater accuracy, validity, and reliability in the development, administration, and interpretation of standardized tests; and

(4) to encourage use of multiple criteria in the grant or denial of any significant educational benefit.

INFORMATION TO TEST SUBJECTS AND POSTSECONDARY EDUCATIONAL INSTITUTIONS

SEC. 3. (a) Each test agency shall provide to any test subject in clear and easily understandable language, along with the registration form for a test, the following information:

(1) The purposes for which the test is constructed and is intended to be used.

(2) The subject matters included on such test and the knowledge and skills which the test purports to measure.

(3) Statements designed to provide information for interpreting the test results, including explanations of the test, and the correlation between test scores and future success in schools and, in the case of tests used for postbaccalaureate admissions, the standard error of measurement and the correlation between test scores and success in the career for which admission is sought.

(4) Statements concerning the effects on and uses of test scores, including—

(A) if the test score is used by itself or with other information to predict future grade point average, the extent, expressed as a percentage, to which the use of this test score improves the accuracy of predicting future grade point average, over and above all other information used; and

(B) a comparison of the average score and percentiles of test subjects by major income groups; and

(C) the extent to which test preparation courses improve test subjects' scores on average, expressed as a percentage.

(5) A description of the form in which test scores will be reported, whether the raw test scores will be altered in any way before being reported to the test

subject, and the manner, if any, the test agency will use the test score (in raw or transformed form) by itself or together with any other information about the test subject to predict in any way the subject's future academic performance for any postsecondary educational institution.

(6) A complete description of any promises or covenants that the test agency makes to the test subject with regard to accuracy of scoring, timely forwarding or score reporting, and privacy of information (including test scores and other information), relating to the test subjects.

(7) The property interests of the test subject in the test results, if any, the duration for which such results will be retained by the test agency, and policies regarding storage, disposal, and future use of test scores.

(8) The time period within which the test subject's test score will be completed and mailed to the test subject and the time period within which such scores will be mailed to test score recipients designated by the test subject.

(9) A description of special services to accommodate handicapped test subjects.

(10) Notice of (A) the information which is available to the test subject under section 5(a)(2), (B) the rights of the test subject under section 6, and (C) the procedure for appeal or review of a test score by the test agency.

(b) Any institution which is a test score recipient shall be provided with the information required by subsection (a). The test agency shall provide such information with respect to any test prior to or coincident with the first reporting of a test score or scores for that test to a recipient institution.

(c) The test agency shall immediately notify the test subject and the institutions designated as test score recipients by the test subject if the test subject's score is delayed ten calendar days beyond the time period stated under subsection (a)(8) of this section.

REPORTS AND STATISTICAL DATA AND OTHER INFORMATION

SEC. 4. (a)(1) In order to further the purposes of this Act, the following information shall be provided to the Secretary by the test agency:

(A) Any study, evaluation, or statistical report pertaining to a test, which a test agency prepares or causes to be prepared, or for which it provides data. Nothing in this paragraph shall require submission of any reports or documents containing information identifiable with any individual test subject. Such information shall be deleted or obliterated prior to submission to the Secretary.

(B) If one test agency develops or produces a test and another test agency sponsors or administers the same test, a copy of their contract for services shall be submitted to the Secretary.

(2) All data, reports, or other documents submitted pursuant to this section will be considered to be records for purposes of section 552(a)(3) of title 5, United States Code.

(b) Within one year of the effective date of this Act, the Secretary shall report to Congress concerning the relationship between the test scores of test subjects and income, race, sex, ethnic, and handicapped status. Such report shall include an evaluation of available data concerning the relationship between test scores and the completion of test preparation courses.

PROMOTING A BETTER UNDERSTANDING OF TESTS

SEC. 5. (a) In order to promote a better understanding of standardized tests and stimulate independent research on such tests, each test agency—

(1) shall, within thirty days after the results of any standardized test are released, file or cause to be filed in the office of the Secretary—

(A) a copy of all test questions used in calculating the test subject's raw score;

(B) the corresponding acceptable answers to those questions; and

(C) all rules for transferring raw scores into those scores reported to the test subject and postsecondary educational institutions together with an explanation of such rules; and

(2) shall, after the test has been filed with the Secretary and upon request of the test subject, send the test subject—

(A) a copy of the test questions used in determining the subject's raw score;

(B) the test subject's individual answer sheet together with a copy of the correct answer sheet to the same test with questions counting toward the test subject's raw score so marked; and

(C) a statement of the raw score used to calculate the scores already sent to the test subject if such request has been made within ninety days of the release of the test score to the test subject.

The test agency may charge a nominal fee for sending out such information requested under paragraph (2) not to exceed the marginal cost of providing the information.

(b) This section shall not apply to any standardized test for which it can be anticipated, on the basis of past experience (as reported under section 7(2) of this Act), will be administered to fewer than five thousand test subjects nationally over a testing year.

(c) Documents submitted to the Secretary pursuant to this section shall be considered to be records for purposes of section 552(a)(3) of title 5, United States Code.

PRIVACY OF TEST SCORES

SEC. 6. The score of any test subject, or any altered or transferred version of the score identifiable with any test subject, shall not be released or disclosed by the test agency to any person, organization, association, corporation, postsecondary educational institution, or governmental agency or subdivision unless specifically authorized by the test subject as a score recipient. A test agency may, however, release all previous scores received by a test subject to any currently designated test score recipient. This section shall not be construed to prohibit release of scores and other information in a form which does not identify the test subject for purposes of research leading to studies and reports primarily concerning the tests themselves.

TESTING COSTS AND FEES TO STUDENTS

SEC. 7. In order to ensure that tests are being offered at a reasonable cost to test subjects, each test agency shall report the following information to the Secretary:

(1) Before March 31, 1983, or within ninety days after it first becomes a test agency, whichever is later, the test agency shall report the closing date of its testing year. Each test agency shall report any change in the closing date of its testing year within ninety days after the change is made.

(2) For each test program, within one hundred and twenty days after the close of the testing year the test agency shall report—

(A) the total number of times the test was taken during the testing year;

(B) the number of test subjects who have taken the test once, who have taken it twice, and who have taken it more than twice during the testing year;

(C) the number of refunds given to individuals who have registered for, but did not take, the test;

(D) the number of test subjects for whom the test fee was waived or reduced;

(E) the total amount of fees received from the test subjects by the test agency for each test program for that test year;

(F) the total amount of revenue received from each test program; and

(G) the expenses to the test agency of the tests, including—

(i) expenses incurred by the test agency for each test program;

(ii) expenses incurred for test development by the test agency for each test program; and

(iii) all expenses which are fixed or can be regarded as overhead expenses and not associated with any test program or with test development;

(3) If a separate fee is charged test subjects for admissions data assembly services or score reporting services, within one hundred and twenty days after the close of the testing year, the test agency shall report—

(A) the number of individuals registering for each admissions data assembly service during the testing year;

(B) the number of individuals registering for each score reporting service during the testing year;

(C) the total amount of revenue received from the individuals by the test agency for each admissions data assembly service or score reporting service during the testing year; and

(D) the expenses to the test agency for each admissions data assembly service or score reporting service during the testing year.

REGULATIONS AND ENFORCEMENT

SEC. 8. (a) The Secretary shall promulgate regulations to implement the provisions of this Act within one hundred and twenty days after the effective date of this Act. The failure of the Secretary to promulgate regulations shall not prevent the provisions of this Act from taking effect.

(b) Any test agency that violates any clause of any provision of this Act shall be liable for a civil penalty not to exceed $2,000 for each violation.

(c) If any provision of this Act shall be declared unconstitutional, invalid, or inapplicable, the other provisions shall remain in effect.

DEFINITIONS

SEC. 9. For purposes of this Act—

(1) the term "admissions data assembly service" means any summary or report of grades, grade point averages, standardized test scores, or any combination of grades and test scores, of an applicant used by any postsecondary educational institution in its admissions process;

(2) the term "Secretary" means the Secretary of Education;

(3) the term "postsecondary educational institution" means any institution providing a course of study beyond the secondary school level and which uses standardized tests as a factor in its admissions process;

(4) the term "score reporting service" means the reporting of a test subject's standardized test score to a test score recipient by a testing agency;

(5) the term "standardized test" or "test" means—

(A) any test that is used, or is required, for the process of selection for admission to postsecondary educational institutions or their programs, or

(B) any test used for preliminary preparation for any test that is used, or is required, for the process of selection for admission to postsecondary educational institutions or their programs,

which affects or is conducted or distributed through any medium of interstate commerce, but such term does not include any test designed solely for nonadmission placement or credit-by-examination or any test developed and administered by an individual school or institution for its own purposes only;

(6) the term "test agency" means any person, organization, association, corporation, partnership, or individual which develops, sponsors, or administers a standardized test;

(7) the term "test preparation course" means any curriculum, course of study, plan of instruction, or method of preparation given for a fee which is specifically designed or constructed to prepare a test subject for, or to improve a test subject's score on, a standardized test;

(8) the term "test program" means all the administrations of a test of the same name during a testing year;

(9) the term "test score" means the value given to the test subject's performance by the test agency on any test, whether reported in numerical, percentile, or any other form;

(10) the term "test score recipient" means any person, organization, association, corporation, postsecondary educational institution, or governmental agency or subdivision to which the test subject requests or designates that a test agency reports his or her score;

(11) the term "test subject" means an individual to whom a test is administered; and

(12) the term "testing year" means the twelve calendar months which the test agency considers either its operational cycle or its fiscal year.

EFFECTIVE DATE

SEC. 10. This Act shall take effect one hundred and eighty days after the date of its enactment.

Chairman **Perkins**. The bill, which was introduced by our colleague, Mr. Weiss, requires that certain information be made available to students that take standardized admission tests.

The Subcommittee on Elementary, Secondary, and Vocational Education conducted many hearings on this bill last Congress. We are delighted to welcome as our first witnesses a panel who will give us an overview of testing: Dr. Sylvia Johnson, professor of research methodology, school of education, Department of Psycho-Educational Studies, Howard University; Dr. William Mehrens, professor of measurement and statistics, School of Education, Michigan State University; George Hanford, president, the College Board; and Bruce Zimmer, executive director, Law School Admission Council.

All of you come around as a panel. We will hear from Dr. Sylvia Johnson first. Mr. Weiss, do you have a statement to make?

Mr. **Weiss**. Thank you, Mr. Chairman. I first want to express my appreciation to both the Subcommittee on Postsecondary Education and the Subcommittee on Elementary, Secondary, and Vocational Education for holding this round of hearings on the Educational Testing Act of 1981. As you know, previous hearings in the 96th Congress yielded a wealth of new information and insight into the standardized testing process. And yet those hearings were successful at really just beginning to make the far-reaching consequences of the standardized testing process known to the American public.

Since hearings last were held a little more than 1 year ago, several events have significantly altered the testing industry. Scoring errors were uncovered as a direct result of disclosure provisions similar to those contained in H.R. 1662—affecting the scores of more than 300,000 students; prominent test-producing and administering agencies—including those we will hear from—have undergone major changes in disclosure policies; and, perhaps most important, public concern over the accuracy and justness of standardized tests and for the glaring accountability gap permitted testing agencies has multiplied among the more than 2½ million students who take these tests every year.

In addition, at least 21 States now are considering Truth-in-Testing legislation which could create a complex regulatory web for test-givers. I am concerned, as I believe all members of both subcommittees here today must be, that the likelihood of a proliferation of State-level disclosure laws may well interfere with standardized testing and thus obstruct our educational process. Federal legislation would improve the testing process by streamlining regulatory demands on test-givers, an issue the testing agencies rightly are concerned about.

Several participants in previous hearings on this issue proposed that we wait and see how successfully the New York State law worked and whether public concern decreased. We have been patient, Mr. Chairman, and today I believe that the fundamental issue addressed by my legislation—making the testing agencies accountable to millions of test-takers every year—is far from resolved.

It is my hope that these hearings will lead to enactment of the Educational Testing Act of 1981. If this legislation is not enacted, we will be continuing to allow testing agencies to operate without making them appropriately accountable to the students they are serving.

Thank you, Mr. Chairman.

Chairman PERKINS. Mr. Peyser, do you have a statement?

Mr. PEYSER. No, I do not, Mr. Chairman.

Chairman PERKINS. I understand that our first witness is not here but Dr. William Mehrens is here. Come around, Dr. Mehrens, take a seat, and go ahead and proceed with your statement. Is George Hanford here? You come around and you be the second witness.

STATEMENT OF WILLIAM MEHRENS, PROFESSOR OF MEASUREMENT AND STATISTICS, SCHOOL OF EDUCATION, DEPARTMENT OF EDUCATIONAL PSYCHOLOGY, MICHIGAN STATE UNIVERSITY

Mr. MEHRENS. Mr. Chairman, members of the two subcommittees, I am pleased to have the opportunity to testify today on the Educational Testing Act of 1981, H.R. 1662.

Let me give you a brief overview of my background. I have been a professor of measurement and evaluation at Michigan State University since 1965. I have published widely in the field including numerous textbooks and articles. I have been active and have held offices in several professional organizations. For example, I have been president of the Association for Measurement and Guidance (AMEG) which is a division of the American Personnel and Guidance Association (APGA). I am currently a senator to APGA and remain on the executive council of AMEG. I am currently on the Board of Directors of the National Council of Measurement in Education (NCME).

My professional activities that are most directly related to my qualifications to testify on H.R. 1662 include: One, working as an NCME board member on their published statement on Educational Admissions Testing, two, helping author the AMEG statement on legislation affecting testing for selection in educational and occupational programs, three, authoring a critical analysis of the Nader report: "The Region of ETA: The Corporation That Makes Up Minds," and four, working as a committee member on the American Psychological Association's standards for educational and psychological test and manuals. I have attached, as appendices to this testimony, the NCME statement and my critical analysis of the Nader report.

I should make clear that my testimony today is mine alone. I am not speaking in any official capacity for AMEG, NCME, nor any other professional measurement organization. However, Bob Linn, current president of NCME, has approved that I submit the attached NCME statement. I perhaps should also state that I have never been employed by a commercial test publisher although I have, on very rare occasions, served as a consultant to a few of them.

When asked if I would testify today I was asked if I would address four questions: One, how are tests developed? two, what do they measure? three how are test results used? and four, what are the positive and negative aspects of test use? To answer these questions fully in 10 to 15 minutes is clearly impossible. The questions perhaps could be adequately addressed in a graduate level seminar in measurement. Nevertheless, I do admire your willingness to become more expert in measurement and I will try to give brief answers to these questions as they relate to the particular type of test referenced in H.R. 1662; that is, any test used for the process of selection for admission to postsecondary education.

How are tests developed and what do they measure?

Test development must actually start by answering the question of how the results are to be used. Since we are discussing tests to be used as selection devices to educational programs, the purpose is to build a test which will assist in the selection decision process. Usually, this means the developers try to develop a test which will predict success in the program. As an aside this is not success on the job. That is a different issue. It is not what the tests ought to be doing but rather to predict success in the program. The test developers determine, through content specialists and reviews of research, what general areas should be tested. These vary from test to test, as they should. For example, the new medical college admissions test covers biology, chemistry, and physics and the content is restricted to areas particularly relevant to medical education. The scholastic aptitude test covers reading comprehension, vocabulary, mathematics, and English, all of which have been proven to relate to success in undergraduate education. We might think of these as achievement tests; for example, the New MCAT, or as measures of scholastic aptitude. The specific terms used are not important as long as we realize that the tests measure developed abilities of the type important for subsequent education success in the program for which the test was developed.

To insure that the tests do this, extensive statistical analyses are done on a large set of developed items and only those which are shown to be useful for the purpose of predicting success are placed in the test. This item/test development process is costly and time consuming. All this is further complicated by the fact that most of the tests of the kind we are speaking of have alternate forms. This requires that the forms be equated so that a person does not have an advantage based on the particular form taken. Further, all test items are looked at carefully and analyzed statistically for potential bias.

This short overview does not do justice to the question of test development. It is a complicated and costly business. It would surely be unfortunate to have to build a test which could be used only once because the items became public information following a test administration as is required by the bill. This would result in poorer tests, complicate the equating process, and increase costs.

How are the test results used? They are used both to help individuals select colleges and colleges to select individual students. If students know their own results and the general distributional characteristics of students enrolled in various schools, they can apply to the schools which are most appropriate for them. The ability of students—along with their counselors—to use this information wisely accounts for the fact that the majority of students are admitted to the college of their first choice.

Various colleges use these scores in quite different fashions. At the undergraduate level many colleges do not require admission exams. Of those that require students to take an exam, some only use it for descriptive purposes to help guide students in their decisions. Other colleges use the examination results along with other data as a basis for admission decisions. I do not know of a single college that bases admission decisions only on tests results.

What are the positive and negative aspects of test use? If one uses test data wisely, better educational decisions will be made. So much data have been gathered to show the validity of college entrance exams for predicting success in college that the finding simply cannot be legitimately disputed.

Tests, because they are objective, are especially useful for those members of society who may be discriminated against if only the more subjective data such as interviews or letters of recommendation were to be used.

There really are no negative aspects of the wise use of test data. Of course, it is possible to overemphasize test data in making a decision but college admissions officers, as a group, do not do this. The error usually goes the other way. Given their proven validity, tests are typically not emphasized enough.

To present more detail in answering the four questions seems not to be the best use of my time. While I cannot make you test experts in 10 minutes, I can give you the beliefs to test experts with respect to the proposed legislation. As mentioned earlier, I am here today to represent my views, I do not formally speak for any organization.

Although speaking for myself I would like to emphasize that I believe my views are quite similar to the views of the majority of the experts in measurement and evaluation. Evidence for this belief can be seen by some brief quotes from the statements of professional measurement associations. For example, the official NCME statement referred to earlier states the following:

"Making all test questions and answers public poses problems for the test developer that may result in increased costs and diminished services for the test taker."

The AMEG statement says, "We do not believe legislative action regarding testing is necessary or desirable." The statement specifically recommends against the enactment of legislation which requires disclosure of items following testing. The statement of the American Psychological Association's Division of Evaluation and Measurement on the 1979 Truth in Testing Act says "Disclosure of items used on standardized tests has many disadvantages." Later the statement says, "We believe that students, themselves, would be the principal losers if the provisions of Section 5 of H.R. 4949 were to become law." That was a reference to the disclosure of items section of H.R. 4949.

I totally agree with the statements quoted from three major measurement organizations. I admit that the assertions made regarding the likely harmful effects of a bill such as H.R. 1662 cannot be proven beyond a shadow of a doubt but these assertions are made by the officers of the three major professional associations in the area of educational and psychological measurement.

They are seriously thought out, informed opinions by the leading experts in the field. You would not take their views lightly.

The stated purpose of the Act—section 2.(b)—are surely not debatable as reasonable goals. However, the bill is not needed for those purposes to be fulfilled and the bill is not likely to facilitate the accomplishment of those purposes, but in fact, is likely to seriously hamper efforts to achieve some of them, for example, greater accuracy, validity, and reliability in testing.

The legislation being proposed is in fact consistent with the efforts by some to do away with, or at least seriously weaken, standardized testing. There is abundant research evidence showing that should these efforts succeed, the result would be poorer educational decisionmaking. In fact, it should be obvious to anyone that better decisions are made with data than without data.

Why do some people in our Nation wish to weaken testing—and therefore the quality of our educational decisions—and why would the proposed legislation have that effect? Some clues to the first part of the question are found in the Nader report referred to earlier. Often the implication is that somehow fewer rejection decisions would be made if tests were not used in selection decisions. Critics of testing appear to believe that if the reason for rejection is low test scores it will be more painful and less understood than if the reasons are low interview ratings, low high school grades, or insufficiently glowing letters of recommendation. Some people even suggest that a positive correlation between income and test score results show the test is unfair.

None of this stands up to a logical analysis. First of all, no college is forced to use a selection test. The testing enterprise is subject to the open market. It is a competitive business. The exception would be the tests produced by the U.S. Government through the Civil Service Commission and the Department of Labor and Defense. Further, the evidence suggests that the majority of students are admitted to the college of their first choice. Tests simply are not ruling lives. Certainly there is no reason at all to believe that, if tests were not used, there would somehow mysteriously be more slots available in college and that therefore one could decrease the number of reject decisions—this number is already quite small for most colleges. The implication that rejection for reasons of low test scores is more painful and less understood than to be rejected for low scores in an interview is simply absurd. Hardly anyone ever understands why he/she received a low score on an interview.

The notion that a positive correlation between income and scores on a test of developed ability indicates the test is unfair is particularly illogical. The whole notion of compensatory education for the disadvantaged is based on the very reasonable notion that parental income is related to education opportunities which are, in turn, related to developed abilities. In fact, the correlation that exists between income and scores such as on the SAT indicates some construct validity for the test. It is supportive evidence for the notion that the test is measuring developed abilities.

I could dwell at length on the mistaken notions of the critics of measurement. They voice many illogical criticisms. They do wish to weaken the testing enterprise and that would lessen the quality of our educational decisionmaking. How would this bill, if passed, serve such an ill-advised purpose?

In brief, it is likely that the bill, of passed, would (1) result in test of inferior quality, (2) increase the costs of testing, (3) decrease services to the students, (4) increase coaching effects—and differentially more so for the wealthy—and (5) result in greater use of locally constructed instruments exempt from the bill in lieu of the better constructed nationally used tests. These points are ones I am sure you have heard before. They are not points which can be proven in advance but they are ones with which I am sure the majority of measurement experts would agree. Let me briefly expand on each of those five points.

The provision of the bill which requires releasing test questions after use will obviously lower the quality of subsequent tests. More test questions will need to be written and used. There will be less research on items before use. The companies will not be able to be so selective in choosing items to put into a test. I recognize that the argument has been made that if there is a sufficiently large pool, a small subset of the released items could be reused. However, most tests do not have a large pool of available backup items of as high a quality as the ones that are placed in the tests and even if a small subset were reused, there would still be an increased demand for more items.

The increase in cost would ensue from the greater number of items needed as well as various other requirements of the bill. No one would possibly urge otherwise. The amount of the cost increase is, of course, open to debate. I do not think it would be trivial. Those who understand item development costs agree with me.

The decrease in service would be an obvious concomitant to the necessity of releasing items. Fewer exam dates would be offered. The result of the LaValle bill in New York documents this.

The release of questions may not, indeed, have too much impact on the effects of coaching. However, if there were an impact it would certainly have to be to increase those effects. This would surely result in a disadvantage to those who could not afford coaching.

Finally, I have suggested that passage of this bill would result in greater use of the locally developed test of inferior quality. This would be true if the quality of the national test and the services provided went down while cost went up. This would be extremely unfortunate for both the individual students and society in general.

As an expert in measurement, I urge you not to consider any so-called truth in testing legislation. It would be a disservice to your constituents.

Thank you very much for your attention.

[Material submitted by William Mehrens follows:]

NCME STATEMENT ON EDUCATIONAL ADMISSIONS TESTING

What Is NCME?

The National Council on Measurement in Education (NCME) is an organization of testing practitioners, scholars, and specialists having as its primary goal the improvement of educational measurement practices and techniques. Its members include college professors who teach and carry on research in the testing field, persons responsible for the conduct of testing programs in state and local educational agencies, individuals in test development and research organizations, and test users. Together, the members constitute a group with well-informed opinions on matters of educational measurement and testing.

What Does NCME Stand For?

NCME regards it as axiomatic that the knowledges, skills, and abilities of people are important and relevant to individual, institutional, and societal goals, and that many of these characteristics are measureable. As corollaries, NCME endorses the following positions:

- Many educational decisions can be improved by objective and equitable assessments of the achievements and characteristics of individuals.
- Tests can provide such assessments which, in many instances, are more objective and more equitable than other forms of evaluation.
- College, graduate, and professional school admissions decisions are among the educational decisions that can be improved by objective and equitable assessment of individual achievements and characteristics to the benefit of the individuals, the institutions, and society.
- The extent to which tests can contribute to sound educational decisions depends on the intrinsic quality of the tests and on proper interpretation and use of test scores.

A major objective of the measurement profession continues to be the assurance of high quality tests and proper test use. Much has been done by the profession to achieve the present high level of test quality and use, and there is every expectation of continuing improvement. This statement, and the proposal which concludes it, is in the spirit of fostering that improvement through the professionals who have the knowledge and expertise about tests and testing, rather than through governmental involvement.

What Does NCME Advocate to Help Test Takers?

> Test takers should be provided with information on the range of knowledge and skills to be tested, the conditions under which a test is to be taken, the types of test questions to be included, and the way a test is to be scored.

This information should be easily available for each test and include, at a minimum, a description of the test, sample questions, a description of the testing conditions, and the scoring procedures to be used. It is desirable to make a complete form of a test or its equivalent available to all test takers. However, to meet this objective it is not necessary to make public all test questions and answers after a test's administration.

> Test takers should know that the test questions which they are called on to answer have not been made public prior to test administration.

If tests were released after administration, and some or all of their contents used again, some test takers would enjoy the opportunity of becoming familiar with the questions and answers. Thus, testing conditions in the future would not be equitable for all examinees, for it is clearly not fair that some, but not all, examinees have opportunity to become familiar in advance with the test content.

Making all test questions and answers public poses problems for the test developer that may result in increased costs and diminished services for the test takers. Two such problems are the maintenance of test quality and score equating.

> Test takers should have assurance that their tests have been scored accurately.

Scoring errors are possible, and consequently procedures to protect test takers from such instances are desirable. Releasing a scored test to the test taker after the testing is only one of several possible ways to provide this protection but is undesirable for the reasons stated above. Alternative procedures to insure correct scoring, such as optional hand scoring services to confirm scores, provide a less disruptive way of accomplishing the objective of assuring a test taker of fair scoring than total disclosure of the test and answer key.

> Test takers, and test users, should have assurance that the tests they are taking, or using, meet professional criteria of good question writing, validity, and reliability.

Measurement experts have developed extensive and rigorous procedures for constructing and validating tests. These procedures are detailed in the *Standards for Educational and Psychological Tests*, developed jointly by NCME and two other professional organizations and relied on in several court cases. Typically, admission test items are written in accordance with detailed content specifications; they are tried out in field studies; and they are reviewed with respect to content, correctness of answers, potential bias, and empirical properties. Reliability and validity studies are conducted and reported. In addition, measurement experts support an independent and extensive review process. For example, the several volumes of the Buros *Mental Measurement Yearbooks* alone contain many thousands of pages of critical test reviews.

Although NCME generally believes that the most widely used admission tests are of high quality, it recognizes its continuing obligation to monitor tests and test use. NCME also recognizes that in the typical professional review, the point of view of the test taker has not always been adequately represented. The conclusion to this statement provides a proposal for monitoring admission testing that emphasizes the rights and concerns of test takers.

What Are Some Issues in the Admissions Testing Context?

Testing has become entangled with a variety of social and personal ideologies, some of which prompt people to argue for changes in educational admissions testing. First, some people object to any evaluation of an individual and, since tests produce some judgments, urge that tests be banned. NCME believes that the value of accurate assessments of relevant characteristics of individuals to the individuals themselves, to society, and to its institutions more than justifies the continued use of tests.

Second, some tests have been viewed as sexually, racially, ethnically, or socio-economically biased because some groups perform less well on the tests than others. Although some interpret group differences in performance as bias in a test, it is now commonly understood among measurement experts that the mere existence of such group differences in performance is not evidence of bias. These differences may reflect individualities of health, growth, and achievement, some of which can be strengthened through increased services and opportunities. Few people suggest that differentials in school-related performances would disappear if there were no tests. Since achievement is desirable for our citizens, NCME believes it is socially constructive to attack the cause of the unequal performances that measuring instruments help discover.

Third, some people feel that ability and past achievement should not be a basis for school admissions and, thus, admission tests are not needed. The issue is whether colleges and universities are to provide universal higher education, or whether some may choose the role of educating the more academically prepared at a more intensive and advanced level. Both roles seem to NCME to be socially desirable, and NCME would not wish to rule out either. Where merit is considered in selection decisions, results from admissions tests can provide objective information about the achievements of individuals in the areas tested.

Fourth, admissions test information has occasionally been misused by individuals in spite of warnings in many test manuals. Misuse includes making admission decisions solely on the basis of test score data, over-interpreting insignificant differences in test performances, and evaluating teachers and schools inappropriately using the admissions test score data. Through its publications and conferences, NCME has committed much of its resources to the effort to reduce to a minimum such improper uses.

What Is Needed?

NCME proposes the establishment of a broadly representative

Commission on the Conduct of Educational
Admissions Testing Programs.

The Commission's membership should reflect the interests of test takers, admissions and other institutional officers, parents, teachers, testing agencies, and the measurement profession.

The Commission should be charged with a single, long-term goal:

> Monitoring and improving the conduct of testing programs whose results are designed to be used in the selection and admissions process in American higher educaion.

Specific activities likely to lead to the realization of this goal are:

1. Developing a clear, workable set of principles that address the concerns of:
   (a) *educators* for reliable and valid informa-

STATEMENT ON EDUCATIONAL ADMISSIONS TESTING, continued

tion for use in the admissions/counseling/advisory processes,

(b) *test takers* for information about tests and the testing process and for an equitable, accurate, and relevant assessment of their knowledge, skills, and abilities, and

(c) *society* for selection procedures that result in an allocation of educational opportunities to a broad spectrum of qualified youth and adults in the population.

2. Seeking voluntary concurrence with these principles by test sponsors and developers and by officials of higher education institutions in the conduct of testing and admissions programs.
3. Establishing a mechanism by which each testing program can be evaluated with respect to the way in which the principles are reflected in its procedures and publications.
4. Reporting the results of these evaluations to the various concerned publics and, if necessary, issuing public censures when instances are detected of repeated divergence from the principles.

NCME Board of Directors
April 1980

This statement has been adopted by the NCME Board of Directors, and can be reproduced and distributed.

EDITOR'S NOTE:

The following article is reprinted in entirety from the April 1981 MEG. *Readers should use this 2nd printing, which corrects production errors contained in the April article, as the definitive version. We regret any inconvenience caused by this inadvertence.*

Criticisms of practices in educational and psychological measurement and evaluation are certainly not anything new. Indeed, the history of these practices shows that such criticisms have been present from the onset of the profession. Professionals and lay persons alike have raised innumerable concerns and issues, in manners ranging from simple questioning to blanket condemnation. In the main, professionals favoring "testing" have accepted these criticisms as challenges and have sought to continue to improve practices so as to assuage criticisms. And so it went, loud attacks and quiet responses.

It now seems that quiet responses are no longer the best tact. Opponents of testing capitalize on modern technology, easily swayed public sentiment, and media prominence to make their criticisms more pungent than ever before. A case in point is the Nairn/Nader report on the practices and policies of Educational Testing Service. Clearly this report demands that alternative perspectives be provided. To ignore this report is to do a grave injustice to those who strive to provide effective educational and psychological measurement and evaluation practices.

For these reasons, I asked William Mehrens to provide an evaluation of the Nairn/Nader report. Bill is well qualified for the task: He has served as president and editor of this journal for AMEG and has authoried numerous books and articles on educational and psychological measurement and evaluation. It should be noted that neither he nor I nor AMEG feels any great need to defend ETS; they can do that on their own. We do, however feel obligated to provide an alternative *professional* perspective. It is in that vein that I commend Bill's article to you. It is important and vital reading. I hope you enjoy it and are stimulated to thought and action.

Larry Loesch

ETS Versus Nairn/Nader: Who Reigns?

WILLIAM MEHRENS

As most *MEG* readers are probably aware, the 554-page Ralph Nader report on the Educational Testing Service authored by Nairn has been published under the title *The Reign of ETS: The Corporation That Makes Up Minds* (1980) (hereinafter referred to as the Nairn/Nader report). The Educational Testing Service (ETS) has responded to portions of this document with two brief reports: *Test Scores and Family Income* (1980a) and *Test Use and Validity* (1980b). In this article, I examine some of the issues, charges, and countercharges in those documents.

If you dislike testing and ETS and

enjoy one-sided reporting, you will like the Nairn/Nader report. If you are a fair-minded scholar who feels positively toward testing and ETS, you will surely be dismayed, as I am, at its biased treatment. Other combinations of characteristics may place you at different places on the pro-con Nairn/Nader report continuum. For example, if you generally believe organizations that are large and successful are also evil, you will view the report differently than if you think largeness and success are characteristics to be admired.

Even if you had heard none of the hype concerning the report and did not pick up a bit of the flavor from the title, the first two sentences in the preface of the Nairn/Nader report (written by Nader) would provide early insight into the direction the report takes: "The conception for this report on the Educational Testing Service began with the *victims* of standardized testing. *Some* of these students would come up to me at colleges and universities around the country to express a *feeling* that they had been *unjustly* judged by a three hour exam" (italics added) (p. ix).

The sentences remind me of a quote by Barclay (1968). He did a delightful job in pointing out that many popular issues are frequently discussed by inappropriate generalizing. He told, for example, of an article written by a Mr. Pulling:

> Mr. Pulling rambled on about testing being similar to phrenology. He said that this article was occasioned by a desultory contact with *some* child who took *some* test in *some* place and was not rated too bright in mathematics. Nevertheless, this particular individual went to *some* college *somewhere* and *somehow* succeeded, all of which prove beyond the shadow of a doubt in Mr. Pulling's logic that *all* testing *in all* places and on *all* levels in similar to the cephalic index of the phrenologists. (Barclay, 1968, p. 4)

The Nairn/Nader report is divided into nine chapters (398 pages), 110 pages of footnotes, and 45 pages of appendixes. It gives the *appearance* of being a scholarly document, but the title, the preface, the chapter titles, and the one-sided set of references soon dispel any such notion. The sections that follow (except for the last two) are responses to the major themes of the nine chapters of the report. In the last two sections I comment on the ETS responses and provide a summary.

PAINFUL UPHEAVALS?

Chapter 1, "Hope . . . Will Be Kept Within Reasonable Bounds," starts out with five case studies of students who supposedly suffered "painful upheavals" caused by ETS. But, the names and *identifying information* have been changed. Even after changing the identifying information, however, the arguments include some contradictions. For example, on page 5, we read that Gary Vladnik "knew that the SAT score meant the end of his college hopes." Then, on page 8, we read that Gary "was lucky enough to live near a good, affordable college which did not require an ETS test." He went there for two years, transferred to the school that had rejected him, and graduated. It is certainly hard to see how this "painful upheaval" resulted in any great harm to Gary, and clearly his ETS score did *not* mean "the end of his college hopes."

This example illustrates two of the underlying themes of the report that disturb me. First, there is a basic theme that ETS has more control over college admissions than it, in fact, has. ETS *does not have any power* over the selection decisions made by the colleges. Colleges do not need to use ETS tests. Many do not. For those that do, ETS does not suggest minimum cutting scores. ETS does *not* tell colleges how much to weigh the SAT scores in the decision-making process and expressly warns them against placing too much emphasis on the SAT or any other single indicator. ETS simply serves as the contractor of the College Boards to build, score, and report the scores on a set of tests. For example, a 1978-79 College Board survey (see Lerner, 1980, p. 128) reported that only 30% of 2,600 colleges surveyed who used the SAT set minimum cut-off scores on the SAT. As mentioned, many other colleges do not even require the SAT. As Lerner points out, "The inevitable overall result is that virtually all literate and numerate students and many semi-literate or even illiterate ones can find some college which will accept them, if they can somehow arrange to pay the fees" (1980, p. 128). Certainly, the evidence suggests that the majority of students are admitted to the college of their first choice. (This would not be quite as true of certain professional schools, such as law schools.)

Second, there is an underlying theme in the report that if students do not get into college, it is unfair and a detriment, both to the individual and to society. At the undergraduate level it is particularly doubtful that not admitting a student to a selective college is of any long-range harm to the individual or society. There are plenty of colleges that will admit anyone with a high school diploma—even if the student is functionally illiterate. If the decision made by a college with admission standards to reject the student is wrong and the student is capable and motivated, he or she will likely do just what Gary Vladnik did: be an academic success. If the decision is correct (and even Nairn/Nader admit the chances are better than 50-50 that it is), the student likely will be better off at a less competitive college.

At the professional school level, such as in law schools, it may be true that qualified candidates get turned away; but with the current surplus of lawyers, that does not harm society. Whether or not it harms the individual is more debatable. If there is a surplus of attorneys, would it be good for an individual to be admitted and graduate from law school if he or she has less aptitude for law than most other lawyers? As Brown (1980) points out:

> To really gauge the importance of the test, it might . . . be useful to ask what happens to people who apply to professional schools and are turned down . . . turned down by the law schools, does the candidate go on to pursue a Ph.D. in history? What percentage of the rejected applicants continue to apply until they are finally selected? What happens to the rest? When all is said and done, how significantly did the test affect the life-chances of individuals who already have bachelor's degrees and are motivated enough to want to go further? (Brown, 1980, pp. 50-51)

A third underlying theme of the report that comes through in the first chapter (although not perhaps illustrated in the Gary Vladnik example) is that somehow rejection for reasons of test scores causes a more painful upheaval than rejection for reasons of low interview ratings, low high school grades, insufficient letters of recommendation from influential advisors, or poor letter-writing skills of the student's high school counselor. The point is, if rejection decisions are made, some individuals will be disappointed. Decisions based on test scores should be no more painful than decisions based on other reasons.

ETS LARGE AND EFFICIENT

Chapter 2 of the report, "Rosedale: Power and Privilege at ETS," is devoted to condemning ETS for its size, efficiency, and concern for quality.

> From around the world, answer sheets are returned to headquarters in Princeton, New Jersey. Propelled by jets of compressed air, these are passed single-file through the entry chute of an electronic scoring machine—the custom-built Westinghouse MRC mark-sensitive scanner. Scanning at the rate of over three sheets per second—24,000 answer sheets per hour—the MRC scores pencil marks representing the answers to millions of multiple choice questions each year. The scores are printed out and forwarded to the mail room where 230,000 letters are processed in an average day. (Nairn, 1980, pp. 32-33)

Taken out of context, we might think the above quote an accolade; but in context, it is obvious Nairn says all this with disapproval.

At another point in the chapter, ETS is criticized for requiring all candidates for the LSAT "to present a photo ID at the test center, give a handwriting sample, sign a pledge that they were, indeed, who they said they were, and submit to thumbprints" (Nairn, 1980, p. 31). Why did ETS do this terrible thing? It was a response "to reports that candidates were hiring impersonators to boost their LSAT scores" (p. 31). Again, it seems most people would praise ETS for this response to curtail cheating. Nairn/Nader condemn them for it!

PREDICTIVE VALIDITY

In Chapter 3, "Five Percent of Nothing: Aptitude Testing, The Respectable Fraud" (and in the accompanying footnotes at the end of the book), Nairn makes some obvious statistical mistakes. The most obvious one is the averaging of the separate validities of the two parts of the SAT, rather than using the validity of the whole test (see footnote 15 to Chapter 3, p. 417). Thus, he averages .37, a validity coefficient for the SAT-V and .32 for the SAT-M and squares this number to get an incorrect coefficient of determination of 11.9. He then incorrectly interprets this as a percentage of perfect prediction and suggests that the SAT predicts grades only 11.9% better than random prediction with a pair of dice (p. 59).

Later in the chapter, Nairn reports that "inclusion of SAT scores in the predictive process improves the prediction of college grades by an average of only five percent or less" (p. 66). He arrives at this figure by computing the coefficient of forcasting efficiency $(1 - \sqrt{1 - r^2})$. For example, for 1974, the index for high school grades alone was 13.4, and for tests and grades combined was 18.5. The difference is, of course, 5.1, but as the ETS response points out, this is a *percent improvement of 38%* (5.1/13.4) (ETS, 1980b, p. 19). Yet, Nairn has made much of his "5% of nothing" claim. Further, as the ETS response points out, in quoting Cronbach and Gleser (1965), the index of forecasting efficiency is not an index that should be used in evaluating tests for selection purposes.

Elsewhere in chapter 3 Nairn suggests that extreme degrees of anxiety are likely to interfere with test performance (p. 86); that the multiple-choice format could favor certain kinds of personalities (p. 88); that paid coaching may improve a person's score and that this raises the question of equity (p. 97); that if a cut-off score is used, a single point can be the difference between acceptance and rejection (p. 156); and that such large decisions about a person's future should not be based on such small samples of performance (p. 159).

Why all these points are made about tests but not interviews or previous grades is puzzling. Certainly, extreme anxiety is likely to interfere with interview performance; an interview may favor certain kinds of personalities; people can be coached to do well on interviews; if a cut-off score is used from *any* type of data, a single point could make a difference between acceptance and rejection; and an interview is a smaller sample of performance (in the sense of time) than a test. But, Nairn prefers not to make these points. Basically, the whole report ignores the limitations of other sources of data for decision-making as well as the extent to which these other measures are currently being used.

HISTORY OF TESTING

Chapter 4, "The Worth of Other Men: The Science of Mental Measurement and the Test of Time," is a somewhat biased review of the history of testing in the United States. Much is made, for example, of the eugenics position ascribed to Galton and Terman, and the use of test data in immigration decisions, quoting Kamin (1974) as follows about the 1924 Immigration Act:

> The law, for which the science of mental testing may claim substantial credit, resulted in the deaths of literally hundreds of thousands of victims of the Nazi biological theorists. (p. 27)

One searches in vain for any evidence that Nairn has read any of the more balanced views of the history of measurement such as that written by Cronbach (1975), who points out that "proponents of testing, from Thomas Jefferson onward, have wanted to open doors for the talented poor, in a system in which doors often are opened by parental wealth and status" (1975, p. 1).

Nairn quotes Thorndike in a disapproving manner when Thorndike suggested that "the able and good should acquire power. In order to support the truth, defend justice and restrain folly, superior men should acquire power" (Nairn, p. 193). The writing does not make it totally clear whether Nairn is opposed to the able and good acquiring power or opposed to the use of test data to help determine the able and good. I rather hope the latter. If the former, then would we want the incompetent and evil to acquire power? I would not. In fact, I would abhor such an occurrence so much that I would want to use as much data as are available (including test data) to decrease the odds of such an occurrence.

CLASS VERSUS MERIT?

Chapter 5, "Class In the Guise of Merit," presents the argument that because SAT scores are related to income, they must not be related to merit—a strange argument. Evidently, Nairn has not heard of sociological studies such as those reported by Havighurst and Neugarten (1975), which show considerable within lifetime and cross-generation changes in social class. Waller (1971) found, for example, that sons who rise above the parent's socioeconomic status (SES) score, on average, better on intelligence tests than sons whose SES is lower than that of their parents. The United States does have a somewhat permeable social class system where financial advancement is based, at least in part, on merit. Thus, one would expect that if the SAT measures merit, and if merit is related to social class, then a correlation between SAT and income should exist. Such a correlation is another bit of data in the network that would support the construct validity of the SAT. To find *no* correlation between the SAT and family income would be evidence against the construct validity of the SAT

Another incorrect implication of the Nairn/Nader report is that tests are perpetuating social classes by keeping poor people out of college. This is simply not true. Fricke (1975, p. 110) demonstrated, for example, that if admission to the freshman class at the University of Michigan had been determined *entirely* by academic aptitude test scores, a *majority* of freshmen would have come from low SES backgrounds rather than the typical 10% to 15% actually admitted. Using only the incorrectly presumed "biased" test scores would not *decrease* but *increase* by a factor of 4 or 5 the number of low-SES students attending the University of Michigan. Of course, no one advocates using only test scores to make decisions. I am only pointing out that it is not low test scores that are keeping low-SES students (in general) out of college.

The ETS response, *Test Scores and Family Income* (ETS, 1980a), makes the following responses to three major assertions made by the Nairn/Nader report.

> 1. *Assertion:* That ETS has attempted to suppress information on the relationship of test scores to students' family incomes.
>
> *Fact:* The data from which Nairn worked is drawn from a series of reports published by the College Board since 1971-72 and distributed to over 15,000 institutions and individuals annually.
>
> 2. *Assertion:* That the relationship of SAT scores to income is inordinately high—that SAT scores and family income rank students in nearly the same way.
>
> *Fact:* While average scores are higher for students from families with higher incomes, students from each income level obtain the full range of SAT scores. Nearly one-third of the students with family incomes below $6,000 rank in the top half of the total group in terms of SAT scores. Other measures of educational achievement show a similar relationship to income.
>
> 3. *Assertion:* That tests are a major instrument in preserving a social status quo, in denying opportunity to students from poor and working class families.
>
> *Fact:* The admissions and financial aid policies and practices of colleges are designed to expand opportunities of low-income students. Use of admissions tests has also contributed to *increased* access of disadvantaged students to higher education. (ETS, 1980a, p. 6)

Another response to the Nairn/Nader report by ETS is the following statement:

> At one point Nairn gives a figure of .40 for the SAT-income correlation, which he states (incorrectly) 'is higher than the correlation which ETS claims to have found between scores and the first year grades the SAT is supposed to predict' (p. 203). He does not indicate that this estimate, which he attributes to Doermann (1968), was based on research using other tests, not the SAT. Nor does Nairn quote the following from Doermann (p. 152): 'In the present study an initial selection of 0.4 was made from among the plausible range of choices suggested by the literature previously noted. While chosen as a best estimate, it also seemed to be a conservative choice for the purposes of this study: that is, of the most plausible choices *it was the highest one* . . .' (italics added)
>
> The typical correlation of SAT scores with college grades (GPA) within colleges conducting validity studies is .41. This correlation is not directly comparable to the SAT income correlation of .29 cited above, since the SAT GPA correlations are computed on selected groups of admitted and enrolled students. The typical correlation of SAT scores and income for these restricted groups is much smaller than the value of .29 found for an unrestricted sample of SAT takers. (ETS, 1980a, p. 8)

This last quote serves as another illustration of the bias in the Nairn/Nader Report.

BARRIER TO THE BAR?

Chapter 6, "ETS: Barrier to the Bar," states that the number of applicants to U.S. law schools exceeds the number of available places by a ratio of 2 to 1, that by 1979 all of the ABA-accredited law schools used the LSAT to screen their applicants, and that the ETS Multistate Bar Examination (MBE) has decreased the importance of the state-prepared essay questions on the bar exam. The report quotes Lerner: "Most American lawyers and judges practicing today would never have gotten into law school at all if they had had to compete against the inflated standards which now govern admission" (Lerner, 1977, p. 6).

If all this is true (and I assume it is), what it means is that we should be able to raise standards in our law schools. Perhaps this is bad, but it seems like a good thing to me. The news reports in Michigan concerning the number of applicants who have failed the bar exams in the past two years do not indicate that our law school graduates are overqualified for their profession. (Incidentally, one report charged that the essay portion of the LSAT was graded too harshly by readers who wanted to limit competition. An attorney who was a reader for the exams responded in part by asserting that he believed it reasonable to expect a law school graduate to be able to spell *defendant*—which many could not do. Whether or not you agree that is a reasonable expectation, this example should indicate that increasing the weight of the essay portions would not be likely to decrease the complaints from those who fail, nor would it be likely to result in fairer, more reliable and valid data.)

Lerner, by the way, objects to the way she is referenced and quoted in the Nairn/Nader report.

> There are six references to my work in the test of the Report . . . and all make it appear as if my analysis of the worth of test scores in selection systems is or was congruent with that of the Naderites. In fact, it is not and never was. The Naderites argue against all uses of standardized test scores in all selection situations; what I argued against was an excessive reliance on test scores to make hairsplit distinctions between high-scoring individuals in hyperselectivity situations, a very different matter.
>
> This is hardly an argument against any use of the LSAT in law school selection, let alone an argument against the use of standardized tests in selection situations generally. And it is most emphatically not an argument for the "new" substitute criteria that the Naderites and their friends are pushing: subjective appraisals, inevitably political, of the worth of candidates' extra-curricular activities and attitudes. (Lerner, 1980, pp. 137-139)

ETS BUSINESS WITH THE STUDENTS

Chapter 7, "The ETS Way of Doing Business: Student Consumers in Captivity," raises concerns about the "monopoly" of ETS in the testing market, the test fees themselves as a barrier to higher education, the contractual relationship between ETS and the test taker, ETS attempts to minimize cheating (Nairn seems to disapprove in this chapter as he did earlier), and the potential for ETS clerical errors in scoring and reporting exam scores. None of these concerns troubles me much, but they might others. ETS does have the best-selling tests in some areas of the market and the only tests in some other markets. They have these because the people who want to use test data require ETS tests and/or contract with ETS to build tests for them. But ETS does not always win against the competition. For example, The American Institute for Research (AIR) was awarded the last contract for the Medical College Admissions Test. My interpretation of the "monopoly"—though not Nairn's—is that, in general, the market place for test purchasing and test use is fairly open. Test publishing is a competitive business. The obvious exceptions are the tests produced by the U.S. Government. Through the Civil Service Commission (Office of Personnel Management) and the Departments of Labor and Defense, the federal government produces its own tests and administers them to about 4.5 million persons annually (see Lerner, 1980, p. 125).

The test fees issue seems minor. Although a 1976 College Board report stated that test fees themselves constitute a barrier to higher education for the very poor, I personally am skeptical. There are a great many costs associated with college attendance. If one can not come up with the money to pay for an entrance exam, it is doubtful that the other costs could be handled either. At any rate, the point is moot because, as Nairn admits, the College Board and ETS have individual fee waivers for needy students. Nairn objects to this system that requires a student to present proof of poverty: "Those reluctant to endure this rite . . . must pay the fee or lose the opportunity to take the test" (Nairn, 1980, p. 262). But, of course, if a student is not willing to admit to poverty, it would also be rather difficult to get a scholarship or any kind of financial aid based on need so the test expense would not really be the delimiting factor.

The business/contractual relationships between ETS and the students trouble me not at all. Nor does the fact that ETS attempts to minimize cheating. The potential for ETS scoring errors, of course, exists. I believe ETS tries hard at quality control, but obviously they are not perfect, and they could try harder.

ETS A TWO-CLASS INSTITUTION?

Chapter 8, "Inside ETS: The Soft Institution," devotes 77 pages to such topics as the nonprofit privileges ETS enjoys, the relationship between ETS and client boards such as The College Board and the Law School Admission Council, and the notion that ETS is a two-class system. No measurement issues come up in this chapter—it is just a general attack on ETS. The attack does take some strange forms. For example, consider the following:

> ETS representatives are flown at company expense to association meetings and conventions around the world when an officer determines that such attendance is justified for professional or business reasons. ETS staff are encouraged to publish papers or books in their fields and hold office in their professional societies. (Nairn, 1980, p. 318)

These are indeed curious "criticisms". Professional organizations have long been advocating such behaviors as desirable professional development and involvement.

The notion that ETS has a two-class system takes up 30 pages of this chapter. What the notion boils down to is that the professionals at ETS are paid more than the clerks. There is a suggestion that this fact disproves the ETS claim that they are like a university! Chauncey and his followers are taken to task for trying "to boost the productivity and decrease the relative numbers of ETS' non-professional clerical staff" (Nairn, 1980, p. 340). Nairn suggests that "non-professionals at ETS have been, *above all*, a cost to be controlled-through speedups, efficiency studies, and careful monitoring of their activities" (p. 347, italics added). Nairn presents no evidence that the "above all" portion of the quote is true and logic alone tells us it cannot be. If that were true, the clerks simply would have been dismissed. Why a concern for efficiency is bad and why the fact that professionals are paid more than clerks is bad escapes my political/social reasoning abilities, although I do recognize certain governmental forms have been advocated by others that would promote inefficiency and equal pay for unequal work.

TRUTH IN TESTING?

Chapter 9, "Rays of Sunlight, Winds of Change," is devoted to promoting so-called truth-in-testing legislation, such as the LaValle Bill passed in New York in 1979. In this final chapter, we are presented with the typical proarguments for such legislation. Rather than present various arguments and counter arguments in this article, I refer the readers to the Educational Commission of the States document *Searching for the Truth About 'Truth In Testing' Legislation* (Brown, 1980). I cannot, however, resist quoting a bit from this chapter. We are told that after the LaValle Bill passed the two Houses, "the testing industry turned its *lobbying* and letter-writing *forces* on Governor Hugh Carey, . . ." but "Ralph Nader *had a conversation with* the Governor" (p. 374, italics added). Ah, but isn't semantics a fun thing?

RESPONSES BY ETS

Test Scores and Family Income and *Test Use and Validity* are short (11 and 27 pages, respectively) responses by ETS to the Nairn/Nader Report. I have already quoted what I consider to be the most important parts of the first document, but I urge the reader to read the rest of it. The second report *Test Use and Validity* points out the same statistical errors that Nairn made and that I have already noted. An introductory note by William Turnbull gives us an accurate overview of the document:

> This paper presents a careful review of the Nairn analysis and finds that in fact it is incorrect: as experience and research have told us for some time, test scores do indeed contribute useful information on which to base admissions decisions. The Nairn exercise manages to arrive at the opposite conclusions through a combination of selective reporting and statistical error, and presents them with a rhetorical flourish that fits the evident intent to demolish the tests rather than enlighten the reader.
>
> It would be a disservice to students, parents, institutions, and the public at large, however, to leave unanswered an attack that marshalls half truths to create an advocate's case against the use of test scores as one element in admissions. This paper is intended to help clarify some of the important issues that may have been confused by the Nader/Nairn report. (ETS, 1980b, p. 5)

SUMMARY

In summary, the Nairn/Nader document is a big, biased blast at testing in general and especially at ETS. Why? Because "the credibility of ETS is essential for its survival" (Nairn, 1980, p. 292). The report attempts to put some chinks in the ETS armor of credibility—or perhaps destroy it completely. But why use such a large weapon? Perhaps because the antitesting forces would like to think of themselves as David and of the testing industry as Goliath. (Of course, David used only a small smooth stone, but then God was on his side.) The analogy doesn't hold, as pointed out by Lerner: The antitesting organizations are indeed quite powerful (Lerner, 1980). The big biased blast will likely be successful in stirring up the emotional forces of the nonscholarly ETS foes. It is not likely to win many converts from among those who favor testing and prefer to make decisions based, at least in part, on dispassionate scholarship. I hope all readers of the Nairn/Nader Report will also read the ETS responses. ETS does not reign in the pejorative way the Nairn/Nader report implies. But, in *writing the truth about testing*, the Nairn/Nader report has reached a nadir, and ETS still reigns if we allow data and logic to rule our minds.

REFERENCES

Barclay, J. R. *Controversial issues in testing* (Guidance Monograph Series III). Boston: Houghton Mifflin, 1968

Brown, R. *Searching for the truth about "truth in testing" legislation.* Denver, Col.: Education Commission of the States, 1980.

Cronbach, L. J. Five decades of public controversy over public testing. *American Psychologist*, 1975, *30*, 1-14.

Cronbach, L. G., & Gleser, G. C. *Psychological Tests and Personnel Decisions.* Urbana: University of Illinois Press, 1965.

Educational Testing Service. *Test scores and family income.* Princeton, N.J.: Author, 1980. (a)

Educational Testing Service. *Test use and validity.* Princeton, N.J.: Author, 1980. (b)

Fricke, B. G. *Grading, testing, standards, and all that.* Ann Arbor: Evaluation and Examinations office, University of Michigan, 1975.

Havighurst, R. J., & Neugarten, B. L. *Society and education* (4th ed.). Boston: Allyn & Bacon, 1975.

Kamin, L. J. *Science and politics of I.Q.* New York: John Wiley & Sons, 1974.

Lerner, B. Equal protection and external screening: Davis, DeFunis, and Bakke. In *Educational Measurement and the Law: Proceedings of the 1977 ETS Invitational Conference.* Princeton: Educational Testing Service, 1977.

Lerner, B. The war on testing: David, Goliath and Gallup. *The Public Interest*, Summer 1980, pp. 119-147.

Nairn, A., & Associates, *The reign of ETS: The corporation that makes up minds.* Washington, D.C.: Authors, 1980.

Waller, J. H. Achievement and social mobility: Relationships among I.Q. score, education, and occupation in two generations. *Social Biology*, 1971, *18*, 252-259.

Chairman **Perkins**. Thank you, Dr. Mehrens. We have now been joined by the distinguished chairman of our postsecondary committee. I understand Dr. Johnson has arrived and you are scheduled to be the next witness. Please take your place at the witness table. You may proceed as you wish. We welcome you.

STATEMENT OF SYLVIA JOHNSON, PROFESSOR OF RESEARCH METHODOLOGY, SCHOOL OF EDUCATION, DEPARTMENT OF PSYCHOEDUCATIONAL STUDIES, HOWARD UNIVERSITY

Ms. **Johnson**. Thank you.

I am glad to have the opportunity to testify before the subcommittee this morning. I will begin by giving you some brief information in terms of my background. I am an associate professor at Howard University in the School of Education in the area of research methodology and statistics. I hold a Ph. D. degree from the University of Iowa. While at the University of Iowa I was special research assistant in testing programs and my degree is in measurement and statistics. Years ago I spent 1 year, a graduate year, at Michigan State University in the program where Dr. Mehrens teaches and I have a masters degree from the area which Mr. Simon represents from Southern Illinois University and undergraduate degree from Howard University. I was asked to make my comments in terms of some areas related to measurements that have some implications for the bill you are considering. Specifically, I was asked to comment in terms of the use of tests for selection and what tests can and cannot do. I would like to begin by considering some points that are perhaps less technical than some of the points that might be raised but I will certainly be glad to go into more technical points if desired.

I think we need to be aware of the nature of measurement and some of the comments I will make here are related to a publication published by the Institute for Study of Educational Policy at Howard University which I wrote a couple of years ago while a senior fellow at the Institute. In terms of the nature of measurement, since we cannot measure minds in the same way we measure the length of a room, all educational and psychological measurement is done indirectly. We never are measuring psychological factors directly. By measuring performance or behavior on some task we infer the level of achievement, intellectual functioning or some other psychological construction. But because measurement is indirect does not mean it is faulty. In fact much of physical measurement is done indirectly. Inferences are made because of certain theories about how the universe operates. These inferences make it possible to measure from distances in the solar system to distances between and within atoms. Physical measurement as well as psychology measurement can result in errors. However, our errors in psychological measurement may be particularly hard to identify and correct. For example, if I say the combination of ingredients used in that wall, if they are mixed together will not harden, and I attempt to test that theory by taking a stroll over into the adjoining office in this direction I will meet with some solid evidence to refute that theory. However, the social scientists lack the benefit of such firm validation or invalidation and is faced with a strong responsibility for careful, thoughtful analysis and empirical validation. In testing the behavior measure used as the basis for an inference is the test score. An inference is then made to a level of achievement, aptitude, personality, interest, or some other human characteristic which cannot be measured directly. The relationship between this test score and the characteristic which we wish to infer may not be the same for all examinees particularly for many members of minority groups who have been systematically excluded from educational benefits. As a result test scores alone frequently are not such measure of aptitude or other psychological constructs for minority groups. Leaders of minority communities have seen numerous cases of discrepancy between ability to perform and measured aptitude. And many regard tests as useless. The reality is that tests are increasingly used in all forms of decisionmaking. Their use and misuses need to be more fully understood by policymakers and community leaders. Many factors operate to attenuate or lower test scores and those factors tend to have the greatest influence on blacks and other minority applicants. These include factors which affect the actual performance of individuals on the tests such as socioeconomic status, differences in educational opportunity, motivation, narrowness of content of the test, atmosphere of the testing situation, and the perceived relevance of the test to success. They also include factors that affect the test score more indirectly such as composition of the group used for item tryouts as well as items selection and item analysis which precede the standardization. The composition of the standardization of groups and technique and procedures are employed in item construction. Also validity and appropriateness of tests often differ for applicants in relation to the same future performance or criterion. These factors attenuate scores and result in validity problems raised in terms of measuring scholastic aptitude.

When we use tests for selection, many tests as you well know are widely used for graduate professional school admissions as well as for admission to colleges. Often people assume that a certain score or certain cutting score automatically eliminates those who are less able and chooses those that are better. We have a problem that arises when it is assumed that if a certain cutting score on a test eliminates potentially unsuccessful applicants, raising that score may result in the better applicant automatically eliminating more of the potentially unsuccessful. In fact this is not always the case. That is many factors need to be considered in terms of college admission rather than a single criterion.

A rough analogy can been seen in the following situation.

Suppose one wishes to board a bus to travel some distance in a large city where there is an established bus fare. Let us assume that fare is a consistent fare for going whatever distance you wish to travel. If you have less than the established fare you will not get a ride. The exact fare is required. Having more money will not take you further under a fixed rate schedule. Whether you reach your destination does not relate to whether you have more than the basic fee in your pocket. It relates to other actions you may take and characteristics you may have, such as how well you check the route map, bus schedule and whether you watch to see if you are nearing your destination. Your timely arrival relates to external factors such as traffic situation, possible accidents, et cetera.

Let us carry our analogy one step further. Imagine an entire college of individuals waiting to board this bus. And some have the required bus fare and others do not. The bus fare will be one method of determining who rides the bus. In this situation there are a whole set of antecedent questions that determine who gets a chance to drop their fare in the box. How widely were the bus schedules and route information distributed? How many people can the bus carry? Who regularly rides this route or knows other people who have taken this bus? Who gets in line first, and who is driving the bus? Once the bus is boarded the same characteristics and external factors determine success in negotiating the system.

An educational example would be helpful here but our image of the test that is the general public's image of the test is as a mental yardstick which makes acceptance difficult. Let us conceive of the test as a yardstick that varies in appearance depending on the purpose for which it is used; a medical school aptitude test used to predict criterion of success as a physician could be viewed as a yardstick with numbers clearly printed in integrals halfway up this stick and blank the rest of the way. The implication—this is a reasonable measure of what we are interested in up to a certain point. After that we need to look at other yardsticks. The same test might be a blank yardstick when used for another purpose such as predicting success as a sculptor or musician. It might be a rubber yardstick if used to predict success in high school science courses. Even though the type of content is appropriate, the inappropriate level would result in measurement with low reliability and validity.

The personal attributes and characteristics that are the focus of many tests are complex entities. Test developers and measurement researchers are aware of the multidimensional nature of intelligence and other aptitudes. Earlier writers noted the general impression that intelligence is unit dimension, is borne out by the common usage of the word. We speak constantly of individuals as having high or low intelligence or having some particular degree of intelligence. This practice is part of our folklore, an example of our tendency to oversimplify complex concepts. Verbal tests with academic type tasks are used to measure academic aptitude, general aptitude and intelligence. The justification of this procedure is that when nonverbal and performance tasks are used they normally could relate highly with verbal testing. And, thus, therefore are more reliable, easy to administer and cheaper verbal tasks are employed. However, the correlations between verbal and performance tasks are not equally high in all groups. They are most likely to be high in those persons who have had the greatest opportunity to develop their verbal skills maximally. Those with the greatest environmental resources for the development of traditional verbal skills. The fact other tasks may more accurately assess aptitude in minorities is often not considered.

So the point we need to consider first, all psychological measurements are not the same. We measure characteristics of people by inference. The same test score does not mean the same thing for everyone. The score of 500 is not always a 500 is not always a 500—

opposite to the rose. Nontest information may give better clues to potential success for low scoring students. However, we must recognize that nontest information can have the same weaknesses as tests in terms of validity and reliability and again must be used carefully. This means the complexity of decisionmaking regarding the future of our young people must be appreciated and dealt with accordingly. There is no simple way to make admissions and placement decisions. Because two groups different substantially on the average on some test, such as SAT, GRE, AMCAT, et cetera, does not mean one group is brighter than the other. It may mean there are factors in the test, the tested and the test taking that result in these differences, typical school and nonschool experiences of one group may better prepare them for the test material, may even mean one group is better prepared, at one point in time for certain educational experiences but it does not mean that people in a lower scoring group are not capable of competent work in the proposed course of study. Even though national legislation regarding tests has not been enacted, its proposal has created a climate within the testing industry that resulted in creative approaches to openness in testing and increased tolerance to those within the industry who have previously advocated use of tests for opening rather than closing options. Tests probably have the greatest utility in providing continual feedback regarding effectivness of educational programing for youngsters. In this use they should be as nonthreatening as possible, taken when the youngster is ready, often by computer, corrections and reasons for wrong answers provided and accompanied by instruction in these problem areas. When used as rigid admissions criteria tests only serve to reinforce the bars to entry and selective programs that have long existed in other ways and they discourage creative way so treating instructional material to make it more accessible to students from nontraditional backgrounds. I would like to thank you for your time this morning and I will be happy to comment further and my written presentation will be submitted.

Chairman. **Perkins.** Your written presentation will be entered in the record without objection.

[The prepared statement of Sylvia Johnson follows:]

Prepared Statement of Sylvia T. Johnson, Ph. D., Department of Psychoeducational Studies, School of Education, Howard University, Washington, D.C.

Good morning Mr. Chairman and committee members. I am glad to have the opportunity to speak with you today regarding the topic of the use of tests in selection for educational programs.

I will begin with a brief account of my background in this area. I have been a faculty member in the Department of Psychoeducational Studies, School of Education at Howard University since 1974. I teach graduate courses in applied statistics, research methodology, and measurement, and I supervise Ph. D. dissertations in educational psychology. I received my Ph. D. from The University of Iowa in Educational Statistics and Measurement. While at Iowa, I was a special research assistant, assigned to the professor who was also the director of Iowa Testing Programs. Several years earlier, I completed a year of advanced graduate study at Michigan State University in the area of Research Design and Development.

I received my masters degree from Southern Illinois University at Carbondale, and completed my undergraduate work in mathematics at Howard University.

Three years ago, I spent a period of time as a senior visiting fellow at the Institute for the Study of Educational Policy, Howard University. While there, I authorized a monograph entitled "The Measurement Mystique." Some of my testimony today is taken from "The Measurement Mystique." However, I do not speak for Howard University, or for the Institute.

When we consider the quality of measurement that we get from tests, it is important to remind ourselves of the nature of the process of measurement.

Since we cannot measure minds in the same way that we measure the length of a room, all educational and psychological measurement is done indirectly. By measuring performance or behavior on some task, we infer the level of achievement, intellecutal functioning, or some other psychological construct.

Leaders in minority communities have seem numerous cases of discrepancy between actual ability to perform and measure aptitude, and many regard tests as useless. The reality is that tests are increasingly used in all forms of decision making, and that their uses and misuses need to be more fully understood by policy makers and community leaders.

How then are tests developed? There are several steps in the process of constructing a test. A test plan is developed which outlines specifications for the test questions, or for sets of questions. A question on a test is called a test item. The specifications may include the type of content the item should contain, the process to be involved in solving the item, the desired difficulty of the item, the format, and any other item characteristics considered important. After a set of items has been written, discussed, and edited, it should be tried out on subjects similar to those for whom the test is intended. This is not the norming, or standardization process, but is an important step, since it helps to determine which items will actually appear on the test. Here the hunches of writers and editors are literally "put to the test", by seeing how well available subjects actually perform.

Often the tryout items are administered in conjunction with an actual previously developed test which serves as a yardstick for judging the quality of the new items. If subjects who do well on the "old" test do well on the tryout item, and subjects who do poorly on the "old" test do poorly on a tryout item, we say that the item discriminated well, or has a high discrimination index. (This is discrimination in a descriptive sense, not in the sense of racial or sexual discrimination). Such an item is said to "work," and is likely to become a part of the test being developed. On the other hand, if subjects who do poorly on the "old" test do well on the tryout item, we say that the item discriminates negatively, and the item is likely to be rejected, revised, or the "correct" answer changed. Often there is something wrong with such items in meaning, style, or some other factor.

Another category of tryout items may occur. These are items that do not discriminate in one direction or the other. That is, subjects who do well on the "old" test perform on a certain level on the tryout item, and subjects who do poorly on the "old" test perform at about the same level. Such items are omitted from the test by many test developers, since they do not appear to add to the information to be gained from the test. These may be called "flat" items, due to the shape of the item characteristic curve, a graph of item performance relative to total score performance.

This group of "flat" items can be quite an amalgam. It may include some poor or misunderstood items, but it may also include items with good discrimination power for subgroups for whom their content is appropriate, and low discrimination power for others. Items with greater validity for minority group members can easily be eliminated in this tryout process, particularly if the proportion of minorities in the tryout group is small. A test that has been well-constructed in some ways, such as careful attention to item format and content, may be seriously biased if the items selected are chosen after tryouts with groups containing few minorities. Some developers do not depend as heavily on tryout data in item selection, but place more emphasis on careful, logical selection by experienced editors and writers. Yet these items writers may be very similar in background and life experiences, as well as in training and academic experiences and thus become an additional source of bias themselves.

Many factors operate to attenuate or lower test scores, and these factors tend to have their greatest effects on Blacks and other minority applicants. These include factors which affect the actual performance of individuals on the test, such as socioeconomic status, differences in educational opportunity, motivation, narrowness of content of the tests, atmosphere of the testing situation, and the perceived relevance of the test to success. They also include factors that affect the test score more indirectly, such as the composition of the group used fo item tryouts and item selection and analysis which preceded the actual standardization, composition of the standardization or normative group, and the techniques and procedures employed in item construction.

In the use of test for selection, it is often assumed by the general public that if a particular cutting score on a test is useful for selection, then raising that cutting score will be even better in selection. In fact, the inclusion for non-test information may give better clues on potential success for lower-scoring students. As a cutting score is raised, many potentially successful applicants are also eliminated, as one seeks to eliminate potentially unsuccessful applicants.

Given a group of qualified applicants, the fact that one student's test scores are somewhat higher than those of another does not mean that the first is more "qualified" than the second. Some basic level of competency should be established, but beyond that many other factors should be used to determine the changes of success.

A rough analogy can be seen in the following situation: Suppose one wishes to board a city bus to travel some distance in a large city, and there is an established bus fare. If you have less than the fare, you will not get a ride. Having more money will not take you farther, under a fixed rate schedule. Whether you reach your destination does not relate to whether you have more than the basic fare in your pocket. It relates to other actions you may take and characteristics which you may have, such as how well you check the route map and bus schedule, and whether you watch to see when you are nearing your destination so that you can head for the exit. Your timely arrival also relates to external factors, such as the traffic situation, possible accidents, etc.

Let us carry our analogy a step further. Imagine an entire collection of individuals waiting to board this bus, some with the required bus fare, and others without it. Then, the bus fare will be one method of determining who rides the bus. In this situation, there are a whole set of antecedent questions that determine who even gets a chance to drop the fare in the box. How widely were the bus schedules and route information distributed? How many people can the bus carry? Who regularly rides this route or knows other people who have taken this bus? Who gets in line first? And, who is driving the bus?

Thus, we can see that a whole set of factors determine success. A good predictor (in this case, the bus fare) can measure some of these factors, but not all of them, and more of one single factor does not necessarily mean that one's chances of success are proportionally increased.

An educational example would be helpful here, but our image of the test as a "mental yardstick" makes acceptance difficult. Imagine the test as a yardstick varying in appearance depending on the purpose for which it is used. A medical school aptitude test used to predict the criterion of success as a physician could be viewed as a yardstick with numbers clearly printed in even intervals nearly halfway up the stick, and then blank the rest of the way. The implication? It's a reasonable measure of what we're interested in, up to a certain point. After that, we need to look at other "yardsticks."

The same test might be a "blank" yardstick when use for another purpose, such as predicting success as a sculptor or musician. It might be a "rubber yardstick" if used to predict success in high school science courses. Even though the type of content is appropriate, the inappropriate difficulty level would result in measurement with low reliability and validity.

The personal attributes and characteristics that are the focus of many tests are complex entities. Test developers and measurement researcher rare aware of the multidimensional nature of intelligence and other aptitudes. Paul Horst, an early measurement researcher noted, "the general impression that intelligence is unidimensional is borne out by the common usage of the word. We speak constantly of individuals as having high or low intelligence or of having some particular degree of intelligence. This practice is part of our folklore and an example of our tendency to oversimplify complex concepts".

Verbal tests with academic-type tasks are normally used to measure academic aptitude, general aptitude, and intelligence. The justification of this procedure is that when non-verbal and performance tasks are used, they normally correlate quite highly with verbal tests, and thus the more reliable, easier to administer, and cheaper verbal tasks are employed. However, the correlations between verbal and performance tasks are not equally high in all groups. They are most likely to be high among these who have had the greatest opportunity to develop their verbal skills maximally, those with the greatest environmental resources for the development of traditional verbal skills. The possibility that other tasks may more accurately assess aptitude in minority youth is not considered.

Now, non-test information can have the same weaknesses as test in terms of validity and reliability, and thus care must be taken in use of these data as well.

The range of problems encountered in using tests in selection does not mean that test information should be thrown out. It does mean that the complexity of decision making regarding the future of our young people must be appreciated, and dealt with accordingly. There is no simple way to make admissions and placement decisions.

Because two groups differ substantially, on the average, on some test such as GRE, MCAT, etc., does not mean that one group is brighter than the other. There are factors in the test, the tested, and the test-taking that result in those differences. It may mean that the typical school and non-school experiences of one group has better prepared them for the test material. It does not mean that people in the lower scoring group are not capable of competent work in the proposed course of study.

In considering legislation related to testing, we must recognize the impact that the proposal of national legislation has had. It has created a climate within the testing industry that has resulted in some creative approaches toward openness in testing, and an increased tolerance of those within the industry who has previously advocated use of tests for opening options to students, rather than closing them.

What, then, are positive uses for tests? Tests probably have their greatest utility in providing continual feedback regarding the effectiveness of educational programming for youngsters. In this use, they should be as non-threatening as possible—taken when the youngster is ready, often by computer, with corrections and reasons for wrong answers provided, and accompanied by instruction in those problem areas.

When used alone as rigid admissions criteria, tests only serve to reinforce the bars to entry in selective programs that have long existed in other ways, and discourage creative approaches to treating instructional material to make it more accessible to students from non-traditional backgrounds.

Comments regarding problem areas in HR 1662-Section 3 Information to test Subjects and Postsecondary Educational Institutions.

P. 4, L. 11-16

A long-range, or conceptual criterion for validating tests used in section may be success in the career field. However, much shorter term validation criteria are employed in practice. For example, little data exists on the MCAT and its correlation with success as a physician, but many studies have correlated the test with first-year medical school grades. One reason for the lack of studies using a long-range criteria is, of course, the time and cost involved. A more important reason, though, is the difficulty in defining and measuring "success" in a field. The fact that is a problem for the test maker can be described clearly to the test taker, along with the reasons, and any proposed plan for research.

P. 4, L. 11-16

This section may seem to give the test taker more information than it actually does. The "percent of variance accounted for" which this statement suggests be given to the test taker is conceptually not useful, and perhaps misleading. There is a possibility for the use of some creative approaches by publishers to convey "the extent . . . to which the use of the test improves the accuracy of predicting future grade point average over and above all other information used". I see an unnecessary limitation in the term "expressed as a percentage".

Chairman PERKINS. I think before we go on to the second panel perhaps we will open this panel up to some questions. Mr. Simon.

Mr. SIMON. First I have to note that Dr. Johnson received her masters degree from Southern Illinois University, which is in my district. She certainly is an expert witness.

Chairman PERKINS. I paid careful attention to that part of her biography.

Mr. SIMON. Let me ask Dr. Mehrens and Dr. Johnson if both of you have had a chance to take a look at the proposed legislation.

Mr. MEHRENS. I have.

Ms. JOHNSON. I have looked at an earlier version of it. I have not had a chance to examine the recent version.

Mr. SIMON. I am advised by the sponsor that it is the same bill. Let me direct this to Dr. Mehrens. Perhaps I should be directing it to the succeeding witnesses and I will be doing that too. Section 3 of the bill which sets forth what you will provide, on page 3 starting line 21, the purpose for which the test is constructed and is intended to be used. That one is no problem. As you go through that series which are the ones that you find objectionable, if any, in that particular section.

Mr. **Mehrens**. It would take me a while to read through them to answer. Statements one and two are not ones which I find objectionable and ones which are already done by all test companies. With respect to interpreting the results statement three including explanation of the test is fine. Correlation between test scores and future success in schools may not be known by the publisher. That would be something the school would have to provide. Those correlations would vary from school to school depending on variety of factors so that is data the publisher may not have available. I would not be opposed to the school providing it if they had that.

Mr. **Simon**. If I may interrupt. I have another question. How many schools, agencies, and publishers are involved in the test scores business?

Mr. **Mehrens**. I would imagine not having sound figures most schools require some tests but not as an admission.

Mr. **Simon**. How many entities originate tests? You have University of Iowa for example.

Mr. **Mehrens**. In terms of test publishers?

Mr. **Simon**. Yes.

Mr. **Mehrens**. There are two major test publishers I suspect that have to do with undergraduate admissions tests, Educational Testings Service and American Council on Testing. There are other publishers such as Psychological Corp., McGraw-Hill, California Test Bureau. That is one. So probably I would think about what I would consider major publishing enterprises.

Mr. **Simon**. Rather than go through in detail, and I would be interested in the response of the next witnesses to the same question, I would be interested in having for the record from you Dr. Mehrens a response to that question.

Mr. **Mehrens**. I will be glad to give you that. If you would like that in writing in a few days.

Mr. **Simon**. Dr. Johnson if I may ask the same of you if you could go through that section 3 and see if there are areas where there are problems and obviously not just to state there is a problem but something of the nature of the problem.

Ms. **Johnson**. I would appreciate the opportunity to go through it thoroughly.

Mr. **Simon**. Then if you can submit that we will enter that in the record.

Thank you, Mr. Chairman.

Chairman **Perkins**. Mr. Erdahl.

Mr. **Erdahl**. Thank you, Mr. Chairman. Thanks to our two witnesses for being with us today and others in the audience that are here as interested participants and listeners. One of the things I would address to both of you if I might, Dr. Mehrens and Dr. Johnson, is that my understanding is—and maybe we should all give Mr. Weiss proper credit that evidently his bill has evoked some modification in policies of testing companies and institutions and if your bill passes or does not pass maybe you have achieved at least part of your goal. As I understand this may be a fair policy in testing so I guess the question I would ask is how does this bill equate as far as need because my understanding is that some testing companies and institutions have already achieved what might be some of the goals or the objectives that are raised in this bill. Would you care to comment on that.

Ms. **Johnson**. As I mentioned in my testimony I think the bill has had the effect of stirring up activity within the testing community. However, it is a bill—in a sense that there can be a lot of activity engendered by the beginning but it is important to have, I think, probably some things in place to continue the impetus of the activity. At the same time I do think that probably there may be some certain aspects of the bill that may in some ways limit test companies in terms of changing test procedures. For example using different ways of administering tests because of certain feedback that has to be given. But I think that could probably be modified. I think that the activity that has been created if it were to continue on its own might not make the legislation necessary. But I do not know whether that would be the case.

Mr. **Erdahl**. It has been my understanding there has been voluntary disclosure by many of the testing companies already. Dr. Mehrens, is that not your awareness?

Mr. **Mehrens**. Yes; I would like to speak on that briefly. Perhaps as a result of the New York bill, perhaps as a result of previous bills introduced at the Federal level there has been more voluntary disclosure of some items. We have an example that has been widely publicized about two scoring errors that were found in a test because of the item disclosure. The inference has been made by proponents of the bill that this will result in better tests. Of course, whether that inference is accurate or not depends upon once we disclose items what we put in place of those items. We will always find faulty items I suspect. I agree very much with Dr. Johnson's position that it is difficult to measure. In the psychological and educational arena the test validation is difficult and to find some items that are faulty and to assume from that that the resultant or subsequent tests will be better is an inference I am not necessarily willing to make. I believe if we expose enough items and we are forced to write new items without the detail we have had in terms of statistical analyses of the current items we may end up with a poorer test. One other brief point with respect to the impact of the proposed legislation—it has stirred up, I think, some forces. One of Nader's goals was to put a chink in the armor of credibility of the Educational Testing Service. I am pretty sure that is not doing a service. That is a disservice. We have abundant evidence that when we compare various ways of making decisions—and again I totally agree with Dr. Johnson we should use a variety of ways—then tests are much more valid than things apt to replace them such as interviews. So stirring up those forces is something I am not convinced will be a good thing for the students in the long run.

Mr. **Erdahl**. Thank you. Another question for Dr. Johnson. How would the bill before us correct some of the inequities that you shared with us? It seems to me that the real problem, if there is one, is the use and application of the test results by the various institutions. If we, as the Congress and as concerned public and as editors and experts, are quite concerned about the results of the tests, that is, some people making poor scores, it seems like we are shooting the messenger that brings the bad news. Maybe there should be some other problem, we should be addressing but it seems to me the real problem is how the test results are used and applied. Obviously they should not be the only criteria for judging whether a person could succeed in some academic pursuit. Would you address yourself to that. I think it is a real problem that you bring to our attention.

Ms. JOHNSON. I think to the extent it promotes a better understanding of tests within the general public it does serve to deal with some of the issues that I raised. That is, if the general public becomes more informed about the limitations of measurement then I think they will impact on the institutions that serve them. In fact, actually the test problem—tests probably are used less within institutions than individuals often think. However, I think many people in the general public think this is somehow skirting or getting around or giving somebody an easy time, whereas I think a greater understanding of what tests can do and how they are made and what their limitations are will make the idea of using a variety of predictors, which in fact many institutions do use, more acceptable in terms of the general public interest. As far as the business of shooting the messenger I always felt that while that may not be appropriate, there is nothing wrong with asking him to sit down and be a little more specific about his message.

Mr. ERDAHL. I think that is a good explanation. One more question, Mr. Chairman, to Dr. Mehrens. You indicated in your testimony, and I think most of us would agree, there is going to be some type of indicators that are going to be used, including testing and interviews by someone. I believe you also gave the impression or maybe you said it directly that the locally originated tests would be not as good or would be deficient. Could you elaborate on that?

Dr. MEHRENS. I would not want to say they would be deficient. The rationale for my opinion on that is simply the sheer difficulties of making these kinds of predictions and the amount of time, effort, and money, and years of analyses it takes to build instruments. It seems to me it is likely that the nationally developed tests which have the amount of money and expertise behind them are going to be greater, resulting in better test in the sense of valid tests for predicting success than one that is developed locally with less expertise, less time, and less money.

Mr. ERDAHL. Thank you very much. Thank you, Mr. Chairman. Mr. Craig.

Mr. CRAIG. Thank you very much, Mr. Chairman. To both Drs. Mehrens and Johnson I would first like to make a brief statement and would appreciate your comment on that statement as it relates to standardized testing. In the information that has been made available to me, and quite a bit of information I have read, it seems the basis for the criticism that H.R. 1662 is attempting to address is the complaint against the standardized testing, and that it discriminates. Now if I were to say that is absolutely correct, it does discriminate. And it would not sell if it did not discriminate or if it did not bring about some, process of measurement. I guess you could liken this to a yardstick. Yardstick measure is feet. You put two objects side by side on the table. One of them is 1 foot long and the other one is 2 feet long. The yardstick becomes the discriminator or the determiner of which one is 1 foot long and which one is 2 feet long. Instead of arguing that both objects are equal and, therefore, something was wrong with the measurement tool, ought we not determine from the tool if those two objects were to be equal? Then, what is the problem in the background or the basis of the two objects, because one is 1 foot long and one is 2 feet long. Is that the basis of the argument in a simplistic way? Are we not using a measurement tool and, in fact, it does discriminate and it was designed to do from a basis?

Ms. JOHNSON. There are relevant and irrelevant bases of discrimination. I think that the problem is that there are factors that can affect test scores that are not relevant to the basis on which we wish to discriminate. When perhaps it might be useful here to comment in terms of the way the test are constructed although Dr. Mehrens has done this to some extent. When you set up a test after a set of items has been written and discussed and edited they are tried out on subjects similar to those for whom the test is intended. I am not talking about the norming process but this is a kind of tryout process. When those items are tried out, it is often done in conjunction with a larger standardized test, a major test that is being given at a particular point in time. There are often tryout items given along with that. Then the scores on those new items are compared with the scores on the bigger standardized test. If subjects who do well on the main tests or this old test we might say, and they do well on the tryout items and people who do poorly on the major test and also do poorly on the tryout items if that occurs, then we say we have good items the test discriminated well. We keep those items. Items that work the other way are said to discriminate negatively and we often toss those out. There is often something wrong with the way they are worded or whatever. Then there may be a variety of flat items. The new item you try out may be the people who do well on the old test and the people who do poorly on the old test do not discriminate between so we toss that out and say it is a flat item. It does not add anything to the score. In fact test items such as that may have discrimination within certain subgroups so that perhaps certain—I am speculating because not that much work has been done in that area—but there may be some items that have particular relevance for some groups that never make it to the main instrument that is being constructed because they, their area of discrimination those with whom, within the groups within which they discriminate may be smaller than the total group. So what you end up with is items that discriminate grossly and you may lose some of the finer points of discrimination. So when you talk about the test being a yardstick, I think you have to realize that yardstick includes both measurements you want and some you do not want. And that has to be recognized. Constancy is a factor in terms of test items. There is an example I use that I would like to share with you. We do not expect tests to ever be content free, of course, but it is these kinds of figures that we try to get test publishers to even out by using contents from this and that and the other but that does not always work. Maybe somebody gets a break because they know something about European history; somebody else gets a break because they know something about art but that does not always work. There was a study done by Medly and Quirk several years ago using national teachers exam and they developed a set of items which had contents of specific interests to blacks even though they developed the items so that the kinds of process that was involved was similar. Sure enough those students did better on the black-oriented items and white students did better on the white-oriented items. Suppose for example you have an item on the GRE that has to do with Great Britain and that knowing something about its history would be helpful. Let us suppose the reading passage actually contains the information that you need to know. Well, someone who knew that before the test has a little break in it even though the content is already there.

Mr. **Craig**. Thank you very much, Dr. Johnson.

Mr. **Mehrens**. I would like to respond briefly also. I certainly agree with the thrust of your statement but I must admit I also agree with the first statement that Dr. Johnson made that there are such things as relevant and irrelevant bases of discrimination. Then she proceeded to give some good examples of the possibilities of irrelevant discrimination. As I read the research evidence on the results of tests for various ethnic groups the various standardized admissions tests, the evidence suggests that they are equally valid across various ethnic groups, that is, the basic correlation between test scores and what they are trying to predict, success in college is pretty much equal. Now this does not suggest that we should not be very alert to the kinds of things Dr. Johnson is raising as possibilities. We should be. It is the fact that, as I view, test companies are alert to these things. Test companies are primarily these days trying to do pretty routinely three kinds of things to be concerned with what we might think of as ethnics bias. One is to have people that represent various ethnic groups and both sexes do a content subjective analysis; that is, look at the items to see if they are apt to be the kind that will give irrelevant discrimination. I think everybody agrees that is not sufficient. That is only a first step in terms of trying to control that. So they also do statistical item analyses with respect to ethnic bias. There are varieties of statistical techniques where I look at the statistical properties of the item within ethnic groups and see if that work on the differential to look at item bias.

Then if item bias does show up in those it is statistically they go back to when the pilot study was done, to the people who took the test and try to find out what their thought processes were or why that bias might exist. The reason I am emphasizing this, and I think it relates to an earlier question, is why I believe there will be those kinds of bias or discrimination in these standardized instruments than there would be in, for example, in interview or locally developed tests that clearly is subject to all the irrelevant discriminations also and does not have the kind of safeguards to steer away from those.

Mr. **Craig**. Thank you both very much. I guess, then, my next question, which I think will bring forth a brief answer is, do you think we can legislate the inaccuracies or the errors that might now exist in some standardized testing methods?

Mr. **Mehrens**. No.

Mr. **Craig**. Dr. Johnson?

Ms. JOHNSON. I think that you can do things through legislation that may assist in that process, not guarantee it.

Mr. CRAIG. Thank you. I would agree. I do not think I can guarantee it because of the various kinds of things that both of you brought up to be very clear as relevant kinds of things that happened within that testing mode.

Dr. Mehrens, one question of you: I understand you recently wrote an article on the Nader report on the reign of ETS. Would you comment briefly on that, especially with their claim that ETS predict grades only slightly better than a role of the dice?

Mr. MEHRENS. Yes; I have included a copy of my article with my written testimony, so it is on file. I do not have that article in front of me. It is in my briefcase but I think I can briefly respond to your comment. They make much of what they call 5 percent of nothing. That is a common phrase they use when they talk about the additional validity of using a test over and above other material. Now, let me just find the right section if I can.

There are some problems; there is not a professional measurement person, so you have some trouble with statistics, but Mr. Nader ends up with an index, to use his figures, of 13.4 for using school grades alone and 18.5, using testing grades combined.

That is a difference of 5.1, but if we look at percentage improvement, I used to teach 8th grade math, but if we go to 18.5, as I taught in 8th grade math, that will be 5.1 over 13.4 or 38 percent improving, using his figures.

He has not even used a percentage increase notion in comparing those two numbers, so even if his numbers were correct, going from 13.4 to 18.5, we would not have a 5-percent improvement but a 38-percent improvement by using the additional data.

It seems to me that there is really no controversy among measurement experts. The combination of high school grades and some sort of aptitude test gives you a better predictor, a significantly better predictor, than using only high school grades alone.

Mr. CRAIG. Thank you both very much.

Let me conclude. In my consultation with some of the universities in my State, one of the student counseling directors, in reference to tests made, comments about some of Mr. Nader's efforts in this whole area. He speaks of the American Personnel and Guidance Association national convention in Chicago a couple of years ago, when he splashed this out on the American scene.

He said that his allegations about testing were so invalid and ridiculous that a fair number of people attending that convention simply got up and walked out. He said his reports and his information offer really no credibility whatever to the issue.

I would like to say, for the record, that my observation is that statement is extremely valid, and I think that the people before us this morning have testified in a broadened sense to demonstrate that there is more to it than just the lack of measurements that Mr. Nader and company are attempting to project into this issue.

Thank you.

Mr. WEISS. Thank you, Mr. Craig.

In the course of your testimony, you stated that the majority of the people in your field, the majority of the experts in measurement evaluation, hold views that are similar to yours.

Now, you are not suggesting by that statement that there are not a great number of people within your field who in fact support openness and disclosure, are you?

Mr. MEHRENS. I suspect it depends in large part about how strict we are in our definition of what constitutes a measurement expert.

Mr. WEISS. People you define as experts; right?

Mr. MEHRENS. I would grant you there is some controversy within the field. My testimony is one I would stick with, and I believe my views are quite similar to the views of the majority of the experts, obviously more than 50 percent, but I would think that must be upwards of 85 percent, if I had to take a guess as to the percent of experts in measurement evaluation who would be opposed to disclosure of items. That is a guess.

I believe my views are similar to the views of the majority.

Mr. WEISS. Well, isn't it a fact that most of the pressure for opening it up, for taking the mystique away from disclosure—because that is really all that this legislation is about, it does not seek to mandate anything of what the contents ought to be or how it is to be used or whatever—that much of that pressure for opening it up came from within the field itself, of people who have been involved throughout their careers in the area of psychometrics, if you will.

Mr. MEHRENS. That is not my impression at all.

Mr. WEISS. Dr. Johnson?

Ms. JOHNSON. Certainly, there has been pressure by individuals within it who have professional backgrounds in measurement to deal in a variety of ways with problems related to measurement, and one of them being legislative, so as far as to where the forces that originated the current movement toward legislation, whether they come from measurement people specifically, I guess it depends a little bit on how you do define measurement people.

I think that there are many minorities who have had some background in measurement and others who have had some background who may not have been active in the professional organizations but who nevertheless maintained an interest in measurement.

Mr. WEISS. Do both of you recognize Dr. Oscar Buros as being an expert in the field?

Mr. MEHRENS. Yes; he is deceased now.

Mr. WEISS. He was an editor of the Mental Measurements Yearbook, one of the winners of the Educational Testing Service's distinguished service awards.

Mr. MEHRENS. Yes.

Mr. WEISS. I give you a quote of his. In 1978, he said:

> It is practically impossible for a competent test technician or test counsellor to make a thorough appraisal to construction, validation, and use of standardized tests because of the limited amount of trustworthy information supplied by the test publishers.

Wouldn't you consider that pressure from within the industry, somebody who is highly expert as to the reason for opening it up?

Ms. JOHNSON. Dr. Burros in the most recent update in the book said that he had hoped by the publication of the Mental Measurements Yearbook there would be some significant improvement in the quality of instruments, and he did not find that to be the case.

Yes; I would agree with you.

Mr. MEHRENS. The quote that you are talking about is one with which I am familiar, of course, and there is nothing in that quote that suggests that we ought to disclose items to the public.

He is concerned about the availability, within your quote, I happen to believe incorrectly, but he is an expert, he is concerned about the availability of evidence regarding the validity of the test to the reviewers. He is in no way suggesting in that quote that we ought to disclose items to the public.

Mr. WEISS. You must be very unhappy with the action of the Educational Testing Service and the College Entrance Examination Board in making 5 out of 12 of their major administrations every year open to the students who request those test answers.

Mr. MEHRENS. Yes; I think that will result in inferior quality.

Mr. WEISS. You are opposed to that policy?

Mr. MEHRENS. Yes.

Mr. WEISS. So long as we know where we are coming from.

You make a statement that you don't know of a single college that based admissions decisions on test results; page 5 of your testimony.

Do you know of any colleges which use the test result as a cutoff?

Mr. MEHRENS. Not as a direct cutoff, I don't, although I have been informed that there are some who have used it in corroboration with other things.

Mr. WEISS. Have you ever looked at a catalog by Lovejoy which lists the various universities and colleges and tells the student and potential applicants that you have to have an SAT score of this amount to be considered for admission?

Mr. MEHRENS. I am familiar with that catalog; yes.

Mr. WEISS. Wouldn't you say that is a cutoff?

Mr. MEHRENS. Surely, that is a cutoff.

Mr. WEISS. I was impressed with your analogy of the bus ride, Dr. Johnson, and the people who have the chance to get on it to begin with, if they have the exact fare, and I wonder if you would comment on the utilization which colleges and universities make of the SAT scores by way of that exact fare provision.

Ms. JOHNSON. Well, if a variety of other predictors are used along with the SAT, then I think you have an improved situation. By the analogy with the bus ride, I meant that if you continually raise the cutting score on the hope that you are getting better and better students, it is really a fallacy.

But if you operate with a certain minimum and then use other factors to determine admission or actions beyond that, you are going to do a fairer job or if you use alternate procedures so that you will examine, if someone should have low test scores, if you examine also their high school grade point average, and the kinds of courses specifically that they had in high school, if one takes a look at some factors that might have entered into the test score and recommend to the students certain programs or certain independent activities that might result in that student being better prepared to take a particular course of study, then you are doing a better service to the students and to the institution than if one is using an arbitrary and increasingly raised score as the number of applicants increases.

Mr. WEISS. I have one final question because time is going on.

Dr. Mehrens, you indicated one of your concerns, although you said you didn't think disclosure would have much of an effect on coaching, but whatever effect it would have would be negative.

You are aware, I am sure, of the recent FTC finding that in fact some of the testing companies had been giving insufficient or inaccurate information as to what the effects of coaching would be on those who took the SAT, that is, the position of ETS had been that there was practically no effect at all of coaching.

The FTC found that that was not so. Now, what is your attitude as to what use those tests should be put to?

Why do you say that it would only benefit the wealthier if you open the tests up at this point, if you made it available to students, test results?

Mr. MEHRENS. That, of course, is not my testimony. What I suggested in my testimony is that it would only benefit them, but to the extent that opening up the test questions will avail the coaching schools of more material to use in their coaching and to the extent that it will be an effect, I don't know how much it is going to benefit those that can afford to go to coaching schools if coaching schools have an impact, I would put the word "binding" in quotes of the FTC study, but certainly they are some of their conclusions; but if you assume that coaching has some impact on test scores, then surely it seems logical to assume that the people who can afford to go to coaching schools will be the beneficiaries of that.

Mr. WEISS. Don't you think one of the effects would be perhaps if you really opened it up that people who otherwise cannot afford to go to coaching school can have local organizations be able to use the return test as a basis for further study?

Mr. MEHRENS. They could. Of course, to the extent that the questions will not be reused, there is no advantage over releasing the questions as opposed to using the current practice prior to the previous practice of having copies of test questions available, so those could have been used all the time.

Mr. WEISS. Dr. Johnson, any comment on the coaching issue?

Ms. JOHNSON. Well, I think that part of the problem in terms of looking at coaching is that one person's coaching is another person's instruction, that is, the idea of what is coaching and what isn't is one of the reasons for conflicts in terms of what has been found.

I think that one thing that coaching can do is to motivate individuals, that is, if kids feel that, hey, that test score isn't the bottom line, in fact, somebody says I can do better if I go and really zero in on this, that may result in improved performance.

That motivation in itself may help to get the student really into learning how the test operates, what attributes of the test taking situation one needs to be in contact with.

I think that there are a lot of things that can be practiced in the sense of getting ready to take a test. A lot of test taking skills, and in a sense almost an emotional readiness, the feeling that one is really able to do it, a lot of that can take place in a coaching or preparation session that can assist a student in doing better.

I would agree that a lot of this is already being done in terms of middle and upper income students, and there is the potential for it being done with lower income students. I don't know whether this will in fact result.

Mr. WEISS. Mr. Simon?

Mr. SIMON. What does coaching cost?

Mr. WEISS. On the basis of prior testimony that we have had, it can run around $300 plus a rental fee for some of the materials, so it can come out to $350 in the most expensive of the schools.

Mr. SIMON. Thank you.

Mr. WEISS. Thank you very much, Dr. Mehrens, Dr. Johnson.

You will submit responses to the questions that Mr. Simon asked you about and we appreciate your taking the time to be with us today.

Our second panel has George Hanford, president of the College Board, and Bruce Zimmer, executive director of the Law Schools Admission Council.

We have your full statements and, without objection, they will be entered into the record and you may proceed as you wish.

You may deliver the entire statement; whatever is your pleasure.

STATEMENT OF GEORGE HANFORD, PRESIDENT, THE COLLEGE BOARD

Mr. HANFORD. My name is George Hanford. I am president of the College Board.

I am a son of a graduate of Carbondale Normal School before it became Southern Illinois University.

I am married to a Minnesotan whose brother is a graduate of St. Olaf's, and I work in New York City.

Mr. ERDAHL. He better stop while he is ahead.

Mr. HANFORD. As you know from previous hearings, the College Board is a national nonprofit association of more than 2,500 schools and colleges committed since 1900 to helping young people get into college.

We are best known for the tests we sponsor, but we provide a lot of other services to help students gain access to higher education. Among them are three testing programs that are administered for us by ETS that would be affected by H.R. 1662.

Because the College Board continues to oppose its provisions and particularly those having to do with test disclosure, I am grateful for the invitation to testify today and to answer the three questions asked in your letter of invitation.

My written testimony makes the case that the College Board has been committed to fair and equitable access to higher education throughout its history.

We have been trying to provide common currencies to make it easier for young people to go to college, academic currencies through tests like the Scholastic Aptitude Test, the SAT, and a currency of financial need through our College Scholarship Service. These common currencies have made it possible for any student anywhere in the country, regardless of status, race, or income to consider any college anywhere in the country, and vice versa.

We oppose H.R. 1662 now particularly because its unlimited disclosure provisions would diminish those possibilities by compromising the utility of the tests used in the admissions process, despite the declared intention of this bill, as we understand it, to promote fairness, equity, and accountability.

In the matter of fairness, our SAT and achievement tests are the most objective assessment instruments of their kind yet devised, free of differences in grading standards among teachers and among schools and free from the taint of personal influence, social position, or family income.

In the matter of equity, we have in my quarter century with the College Board seen the proportion of minority youth in higher education grow from 3 percent to about 16 percent. And as for accountability we have seen the number of young people in higher education increase from 2½ to 12 million, a remarkable record in itself and all the more so when compared to that of any other country in the world. I see no reason for this kind of governmental intervention in a process that has succeeded this well without it.

Now to the three issues you have asked me to address:

First, the reasons why the College Board changed its policy with respect to disclosure. There are essentially three: First, to be sure, we have changed our practice with respect to test disclosure but we haven't changed our position with respect to openness.

When I joined the College Board staff in 1955 high school seniors were not allowed to know their SAT scores. But by the time my older daughter took the SAT in 1962 we were not only reporting scores but also making half an SAT available for practice.

By 1978 we were publishing "Taking the SAT," a full practice form. Disclosure, in this evolving context, is but a natural extension of the process of opening up the college admissions testing process that has been going on since World War II, an extension obviously accelerated, I am the first to admit, by the spotlight of public attention put on the process by this controversial truth-in-testing legislation.

Second was the board's desire to demonstrate again that it is both sensitive and responsive to public interest in the testing process, and that it is possible for the board, given the time to do so, to develop an appropriate program of action, even where we may question the educational value involved.

We have by our action proved that the self-regulating and voluntary system, given the time, can address legitimate concerns without this kind of governmental intervention.

The third reason for the change has to do with the New York experience. When the New York Act was being considered we argued not only that it was unwarranted but that it was also too hastily conceived. A series of amendments have had to be enacted to deal with some of the more troublesome features of the legislation; but other problems remain.

When similar bills were subsequently introduced in 25 other States and in this Congress we argued, "Let's wait and see how the New York law works." Our advice has been heeded; so far no other bills have been enacted.

We carefully and deliberately studied what was going on in New York. Most importantly we found out that not many candidates wanted the disclosure service. No more than 3 percent of the students who took the SAT in New York in 1980–81 took advantage of our question-and-answer service.

We concluded that, at that level of demand, five national administrations which now accommodate more than 70 percent of our candidates would effectively meet the wishes of any student anywhere who would want to utilize the service.

We also used the opportunity to confirm that we could indeed live with the disclosure of five SAT editions a year without compromising the convenience, the quality, or the fairness of the tests, and so we have chosen to disclose five editions nationally to provide limited, I stress limited, disclosure with full service nationally, while we are forced by the unlimited provisions of the New York law to curtail services and increase prices there.

The board's continuing commitment to openness in the college admissions testing process, our demonstrated responsiveness to public concerns, and our good faith follow-through on our expressed intention to act in the light of experience with disclosure in New York demonstrate, I believe, Messrs. Chairmen and members of the subcommittees, that the board is not at odds in principle with the stated objectives of these bills.

And I should say here that however much we may differ on the substance, legislators like yourselves, concerned about education, have performed an important service by raising the issues which we are addressing in this forum.

At the College Board, I can assure that that we are doing all we can at this point in time. If we are asked to publish each and every SAT edition after one-time use, the quality and the fairness of the tests will be compromised and the opportunities to take them, particularly on the part of the handicapped and Saturday Sabbath observers, will be greatly diminished if not destroyed.

And if H.R. 1662 were to apply, as is proposed, to our subject matter achievement tests, the character of those offerings is such that we would undoubtedly, albeit most reluctantly, be forced to discontinue them. I doubt that you really want to do such damage to a process that has served so many so well.

The second and third matters which you have asked me to discuss relate to the two flawed test questions which received so much attention during the past year. The discovery of one was the result of a voluntary service of the College Board which is to return PSAT forms free to all students nationally. The other, of legislatively mandated disclosure of the SAT in New York State. Both questions were found to have, among the answers presented, one that was correct in addition to the one that was intended as correct.

The meaning of these flawed questions is that the tests, like all human artifacts, are subject to human fallibility. In one sense, the wonder is that anyone would be surprised that a mistake can occur on a College Board test.

In another it is that out of more than 1,500 questions placed under scrutiny by students and the public as a result of our various publications since January of 1980 only these two were found to be flawed.

In both of these instances, where scoring was already completed and reported, the Educational Testing Service at our direction rescored all of the tests, giving credit for both correct answers and rereporting the scores to all receiving institutions even though the difference in any individual SAT case was no more than 20 points on the test's 200–800 scale.

In other words, even when differences are small our policy is to correct them, not only to avoid any possible unfair consequences to the students but because to do otherwise would be to compromise the credibility of a process which we view as a public trust.

Although our policy for many years has been to entertain and consider seriously any challenge to the accuracy of a test score, or the soundness of a test question, at any point, and when appropriate to take remedial action, we have recently formalized and expanded the procedures for doing so. Where the views of a challenger, generally a student, and the view of Educational Testing Service cannot be reconciled, we have added a procedure for independent, impartial third-party review and determination.

I believe this as well as other actions I have described to you demonstrate my basic point. The College Board was founded under voluntary educational auspices and functions as a public trust to perform a variety of services to help students and institutions make better decisions about each other and themselves. We do not make those decisions. The tests we sponsor are means, rather than ends, means to help implement a process which is profoundly rooted in the American way of doing things.

We intend to keep those means responsive to the interests of students, institutions, and the public. They are part of a system which is certainly susceptible of improvement but which works better than any other in the world, is more consistent with our national values than any other could be, and should be given the opportunity to continue its pursuit of equal and expanded opportunity, openness, fairness and accountability without the burden of the kind of governmental regulation proposed in H.R. 1662.

May I say simply in conclusion that I look forward to the possibility of returning to Washington later this year to discuss with Mr. Perkins' subcommittee the pressing matter of upgrading the quality of schooling in America.

H.R. 1662 focuses on the role of tests in the educational enterprise, but a much more fundamental public concern today has to do with the performance of our schools. For the eighties the College Board initiated a decade-long project, Project EQuality, for the purpose of helping to direct attention to basic academic skills and raise educational standards.

I would welcome the opportunity to share with the subcommittee more about our plans and objectives in this effort.

[The prepared statement of George Hanford follows:]

PREPARED STATEMENT OF GEORGE H. HANFORD, PRESIDENT, THE COLLEGE BOARD

Messrs. Chairmen and members of the Subcommittees, my name is George H. Hanford. I am president of the College Board.

The College Board is a not-for-profit national association of more than 2,500 schools and colleges, both public and private in their governance. It was founded eighty-one years ago for the primary purpose of making the process of getting admitted to college easier and fairer for more students than had previously been the case; in other words to expand access to postsecondary education. This has continued to be its purpose.

The Board provides services to students an institutions in college admissions, student financial aid, counseling, guidance, placement, and credit by examination. These services are sponsored by the members in furtherance of the Board's purpose, but are available to all individuals and educational institutions that wish to make use of them. The policies under which they are provided are established, and the services themselves are continuously overseen, by a Board of Trustees elected by the membership.

Among those services are certain standardized test, including the Preliminary Scholastic Aptitude Tests/National Merit Qualifying Test (PSAT/NMSQT), the Scholastic Aptitude Test (SAT), and various Achievement Tests, which would be affected by HR 1662, the Educational Testing Act of 1981.

The Board contracts with Educational Testing Service of Princeton, New Jersey to prepare and administer these tests. But as sponsor and representative of the institutions which use the tests, the Board set the policy and is responsible for the process.

I am therefore grateful for the invitation of the Subcommittees to testify on HR 1662, the Educational Testing Act of 1981. The invitation specified three questions which the Subcommittees wished me to address. I shall do so in the context of a brief general statement.

The declared purpose of these bills is, among other things, to promote fairness, equity, and accountability in the use of tests in college admissions by giving students and the public more information about these tests.

The issues of fairness, equity and accountability have been primary concerns of the members and staff of the College Board from its inception. Our tests themselves—first only in academic subject matter areas, later joined by standardized tests of developed mathematical and verbal reasoning ability—were introduced at the outset as one reliable means of establishing a "common currency" of academic attainment by college-bound students no matter where they came from or where they wanted to go to college. They afforded one of the few assurances of objective and uniform assessment that our educational system could offer to document mastery of the skills and knowledge that were—and are—considered essential to success in doing college work.

These same tests, because of their uniformity and objectivity, their freedom from the taint of personal influence or social or economic position, have made it possible, especially in the years since World War II, for millions of high school graduates who might not otherwise have been considered "college material" to validate their qualifications for such opportunity and gain admission to college. They have been an indispensable tool in advancing the cause of equal as well as expanded educational opportunity.

The Board's concern with equal and expanded opportunity has not, however, been confined to testing as such. In the 1950s, the Board pioneered the concept of need-based student financial aid, thereby helping to open the door to college for millions of economically disadvantaged students. It developed, and today conducts, the nation's largest program to help students present their economic needs to colleges, and to help colleges work out the means to help them through scholarships, loans and jobs to pursue their educational aspirations. We see this service as an essential complement to the service of helping students to present their academic credentials.

In the early 1960s, the Board moved to make those services equally and inseparably available to minority and majority students alike by desegregating all of its testing centers throughout the nation—often at the risk of personal harm to those involved. But it was done, and even today testing sites are selected with attention to minimizing the sense for any student that he or she is entering foreign territory to take the tests.

The cause of fuller information for students and better public understanding of tests and the admissions process (including the intricacies of student financial aid) has also been the Board's cause. Again in the 1950s when the conventional wisdom in education said that students should not be given their scores on standardized tests because they might not understand them properly, the Board adopted the practice of reporting scores directly to students as well as to their schools and the colleges to which they were applying.

The standard practice of presenting sample questions and a full explanation of the nature and form of the tests was adopted in the Board's earliest years. And the

Board has consistently acted to expand the information provided to students as advancing knowledge demonstrated the educational soundness of doing so, and the interests of students indicated its desirability. The Preliminary Scholastic Aptitude Test was devised as a means of providing guidance information earlier in their high school careers. Practice samples of the SAT were offered in the 1960s and in 1978 the Board began providing free to all registrants a full sample SAT—an actual test—which could be taken under timed conditions for further familiarization and guidance.

All of which means, Messrs. Chairmen and members of the Subcommittees, that the Board is not at odds in principle with the stated objectives of this and similar bills. And I should say here that however we may differ on the substance, legislators concerned about education have performed an important service by raising the issues which we are addressing in this forum. The broad public discussion that has been generated in these halls and in state legislatures across the nation has helped us to create broader public awareness of how admissions tests are used—to help students get into college rather than turn them away—than our educational efforts could have achieved alone.

Nevertheless, we oppose such bills. Not because we oppose the interest and concern they express, which we share, but because we believe that the Board's educational commitment and actions, and those of other test sponsors, make these bills unnecessary and therefore inherently undesirable; and because the specific provisions of these bills would adversely affect our ability to sustain the present level and quality of service which we offer to students and institutions. Other test sponsors will surely speak for themselves, but I believe it is fair to say that all of them are committed, as is the Board, to providing leadership in promoting fairness, equity and accountability, and to the furtherance of public information and understanding of testing and its place in the process of admission to higher education.

The college admissions process no less than curriculum content is a sensitive educational area which is uniquely carried out in this country by purely voluntary and pluralistic means, consistent with the deepest values of our democratic system. We feel that standardized testing and the admissions process of which it is a part should remain the responsibility of the educational community, in a voluntary and pluralistic system nurtured by the schools and colleges themselves without federal or state intervention. This country is blessed with perhaps the best and most egalitarian system of higher education in the world, and the system that created it should not be altered without compelling reasons. We do not think such reason exists.

Let me now turn to the first of the Subcommittees' questions: the reasons that the College Board adopted a policy in March of 1981 to provide a national question-and-answer service at five administrations of the Scholastic Aptitude Test, so that any student at such administrations could order copies of his or her answer sheet, the correct answers, and a copy of the questions that counted toward his or her score.

I have said that the Board historically has pursued a policy of opening information about the tests to students and the public, and I have highlighted some of the earlier actions along this line. I should also note that just in the last year and a half, the Board has substantially increased its information services in several other ways: first, by providing a similar question-and-answer service to all students taking the PSAT/NMSQT; by providing a national score verification service option for students who wanted to check their scores but did not want all of the questions and answers; and by publishing and distribututing widely copies of additional actual tests, including one accompanied by full technical data on its development and performance in use.

In this context, the March action is an extension of policies already in effect, and it was taken for three main reasons:

The first reason was, in fact, our standing policy of commitment to full information: to expending information about tests to students and the public so long as it does not adversely affect the cost or quality of services to students, and/or the reliability and fairness of the tests.

The second reason was the conclusion, reached by our Board of Trustees after careful study of our experience with a New York State law similar to HR 1662 that 1) such service and information could be offered on a *limited* basis to all students without damaging the quality of the test or disadvantaging students who need special administrations, and 2) that the very small demand demonstrated in New York State for the question and answer service—the overall total for the 1980-81 testing year was 3 percent—meant that to offer the service nationally at administrations which now accommodate 70 percent of students would effectively meet the wishes of any student anywhere who would be likely to want the service.

The third was the Board's desire to demonstrate again that it is both sensitive and responsive to public interest in the testing process, and that it is possible for the Board, given the time to do so, to develop a program of action that demonstrates such responsiveness—even where we may question whether there is educational value to be gained. Our goal in the words of the Board's chairman, Dean Fred Hargadon of Stanford University, was to "satisfy both proponents of test disclsoure *and* proponents of a testing program that produces reliable and valid results." In short, we sought to prove that the self-regulating and voluntary system, given the time and opportunity, can address legitimate concerns without the penalities that accompany this kind of governmental intervention.

We believe we have done that. Although offering very extensive opportunities for the question-and-answer service, the schedule adopted by the Trustees does not sacrifice those additional opportunities to take the test that afford convenience to some students and institutions, and meet the special needs of many more of them.

I must also tell you in all good faith that we are at this point doing as much as the current state of the testing art permits in our program. If the Board were required to provide question-and-answer service at additional administrations of the SAT, the character of our program is such that we would be forced to reduce our schedule to those administrations where we now plan to offer the question-and-answer feature, and to eliminate all other regular administrations, and most special administrations for institutional purposes, makeup, and to meet the needs of handicapped students.

Furthermore, if the law were to apply to our subject-matter Achievement Tests, the character of that program is such that we would undoubtedly, albeit most reluctantly, be forced to phase them out.

The second and third questions which the Subcommittees posed had to do with the discovery of two flawed test questions, one in a nationally administered edition of the PSAT/NMSQT, and another on a form of the SAT administered in New York State. The discovery of one of these was therefore the result of voluntary service of the Board, in returning PSAT forms to all students, and the other the result of a legislatively mandated service in New York State. Each question was found to have, among the answers presented, one that was correct in addition to the one that was intended as the correct answer.

The meaning of these flawed questions is that the tests, like all human artifacts, are subject to human fallibility. In one sense, the wonder is that anyone would be surprised that a mistake can occur on a College Board test. In another, it is that out of more than 1,500 questions placed under scrutiny by students and the public as a result of our various publications since January 1980, only these two were found to be flawed.

We are not complacent about such flaws, and we try to avoid them through extensive quality control. Test authors, editors, and test development committees of high school and college people all review the questions repeatedly and intensively for flaws of all kinds—including content bias, ethnic bias, and ambiguity.

This scrutiny continues beyond the administration and scoring of the test, and in those rare instances when a flaw is discovered by us or anyone else at that point, our policy is very simple. We correct the error with a primary regard to fairness for all students. In both recent instances, where scoring was already completed and reported, the Educational Testing Service at our direction rescored all of the tests, giving credit for both correct answers, and re-reported the scores to all receiving institutions even though the difference in any individual case was no more than 3 points on the PSATs 20-80 scale and 20 points on the SATs 200-800 scale. In other words, even where differences are small, we correct them not only to avoid any possible unfair consequences to the students, but because to do otherwise would be to compromise the credibility of a process which we view as a public trust.

Although our policy for many years has been to consider seriously any challenge to the accuracy of a test score or the content of a test question at any point, and where appropriate to take remedial action, we have recently formalized and expanded the procedures for doing so. Where the view of a challenger—generally a student—and the view of Educational Testing Service cannot be reconciled, we have added a procedure for independent, impartial third-party review and determination.

I believe this and the other actions I have described to you demonstrate my basic point. The College Board was founded under voluntary, educational auspices, and functions as a public trust to perform a variety of services to help students and institutions make better decisions about each other and about themselves. We do not make those decisions. The tests we sponsor are means, rather than ends, to implement a process which is profoundly rooted in the American way of doing things. We intend to keep those means responsive to the interests of students, institutions, and the public. They are part of a system which is certainly susceptible of improvement, but which works better than any other in the world, is more consistent with our national values than any other could be, and should be given the opportunity to continue its pursuit of equal and expanded opportunity, openness, fairness and accountability without the burden of the kind of government regulation proposed in HR 1662.

May I say, finally, that I look forward to the possibility of returning to Washington later this year to discuss with Mr. Perkins' Subcommittee the pressing matter of upgrading the quality of schooling in America. HR 1662 focuses on the role of tests in the educational enterprise, but a much more fundamental public concern today had to do with the performance of our schools. For 1980s the College Board has initiated a decade-long project—Project EQuality—for the purpose of helping to direct attention to basic academic skills and raise educational standards. I would welcome the opportunity to share with the Subcommittee more about our plans and objectives in this effort.

Mr. WEISS. Thank you very much.

Mr. Zimmer?

STATEMENT OF BRUCE ZIMMER, EXECUTIVE DIRECTOR, LAW SCHOOL ADMISSION COUNCIL

Mr. ZIMMER. My name is Bruce Zimmer, executive director of Law School Admission Council, a nonprofit educational association whose members are all of the ABA approved law schools.

Directly and through our operating corporation, Law School Admissions Services, we provide a wide array of services and publications related to law school admission.

One of those is the Law School Data Assembly Service, the national system of gathering and distributing academic and biographical information for the law school admissions process.

We put a substantial effort in this program in order to render more readily comparable and democratic records and evaluation of applicant academic records in the admissions process.

Each year information concerning 100,000 people from 3,000 colleges applying to over 250 law schools is collected, processed and communicated through this system. The law school admission test is another valuable component of the national law school admission program. We contract with the American college testing program and Educational Testing Service for the development of the law school admission test.

The test is administered under the auspices of our operating corporation, Law School Admission Services.

The Law School Admission Council does not support added legislation to regulate educational testing agencies or test use. We are not persuaded that there remain problems which demand legislative remedy, or that the asserted benefits of such legislation would outweigh the detriments of additional regulation of an educational community already burdened with extensive regulation.

Further, the impact of voluntary and responsive actions of testing organizations is steadily becoming more apparent to our consuming and evaluating publics, and especially to our critics. The recent record of test programs like ours may or may not document a need for some forms of regulation of educational testing. We welcome study of this question.

We support the continued oversight of Congress in this matter. But we do not think a case for Federal admission test legislation has as yet been made.

The idea of Federal testing legislation is not in itself unthinkable. Under certain potential circumstances a national law could preempt conflicting, confusing and oppressive State enactments. One appropriate and coherent scheme of regulation would be preferable to many uncoordinated and burdensome schemes.

Indeed, our own research has led us to conclude that a sound model preemptive act could readily be developed. But the circumstances we face today suggest that enactment of a Federal test law in the near future would be premature and probably unnecessary.

The national debate on testing for higher education admission has come a very long way in the 4 or 5 years that testing bills have been receiving serious attention from State legislatures and from Members of Congress and their staffs. The environment in which testing programs operate has changed dramatically since July 1979 when New York enacted a test disclosure law to become effective at the start of 1980.

At different rates, many testing organizations have made major, indeed what 2 or 3 years ago would have been thought revolutionary, commitments to openness in their testing programs.

The Law School Admission Council has offered test disclosure for all of our tests throughout the United States and throughout the world, other than some very limited special administrations, since the New York law took effect.

The New York test law did not permit us to continue our historic method of making scores comparable from test to test, and still disclose every form. Our decision to adopt universal test disclosure thus triggered a significantly revised structure for our test.

As a result of that change we are planning to introduce a revised LSAT in June 1982. It will be comprised of four scored reading and reasoning sections, two unscored pretest or equating sections and an unscored writing exercise for individual law school evaluation.

A 10 to 50 score scale will replace the current 200 to 800 scale. We have put into the record some substantial explanatory material.

Mr. WEISS. Without objection, that will be received.

Mr. ZIMMER. We are presently contracting to develop an adequate inventory of this education of the LSAT. Having developed such a supply, we plan to address ourselves to emerging methods and technologies in evaluation in order to provide to law school applicants in the late eighties and early nineties the best academic measure that can be provided at that time.

Test disclosure is a service law school test-takers desire, and for which they are willing to pay. Because consumer demand for disclosed tests had become so substantial, about 60 percent of our test-takers, by the second year of test disclosure, and because of an enhanced interests on our part in providing an extra increment of quality assurance for our test, the Law School Admission Council Board of Trustees elected to make test and answer record disclosure an automatic part of the reporting procedures for test-takers, beginning with the 1981-82 test year.

This is a culmination of basic commitments on our part. We are rebuilding our services to accommodate our consumers' interest in openness in testing.

Two years ago the State of New York took a decisive position on test disclosure. It insisted on immediate action on full test disclosure. A sensible scheme of phased applicability would have been more fruitful. By giving 4 or 5 months notice for radical change to organizations that had hitherto operated on much longer test development leadtimes, New York disrupted the orderly conduct of testing programs like ours.

Its actions caused increases in applicant fees that probably would not have been undertaken on the initiative of testing organizations alone. The 1979 New York law produced substantial technical problems and some areas of unintended encumbrance and intrusion. Several of these problem areas were modified in 1980 in a cooperative effort in which our organization took an active part.

Over time full analysis of the impact of disclosed testing on test results must await research that can be performed only after several years of disclosed testing.

In no small measure the New York act has worked reasonably because of the prudent regulations and sensible enforcement procedures of the New York Education Commissioner and his staff.

Indeed, as the various test programs adjust to disclosed testing, there is a greater risk that ill-advised regulations and enforcement procedures will emanate from the State educational bureaucracies than there is a probability that substantive provisions of a New York-type law would, themselves, prove intolerant if enacted in another State.

There is little question that to date New York has effectively preempted the national testing issue. It has acted, in effect, to set a national standard for many test programs. As Pennsylvania has done in the area of agricultural product standards and as New York itself has done in the insurance law field, it has established a de facto national standard.

New York has done this by providing a powerful impetus to increased test disclosure nationwide. That impetus has come by virtue of the need and desire of test sponsors to maintain testing opportunities for New York students and schools, as well as from the momentum of public opinion arising as test disclosure increased scrutiny of the underlying soundness of admission testing.

Testing organizations like ours have discovered that, on balance, test disclosure promoted public confidence in the integrity of the test instrument and the testing process. It became symbolic of our public accountability. Indeed, openness has helped us in our quality control efforts.

As other States have considered test legislation in the last 2 years, nearly all have deferred to the New York law and experience as reference points. Most have concluded that the considerable progress toward national test disclosure made by programs like ours, steadily joined by a growing list of major test sponsors, had obviated the need for State-by-State regulation.

In many instances, test sponsors have accelerated their commitment to openness in order to show good faith and thus eliminate a need for potentially burdensome regulation. During this past year's sessions there was support in several State legislatures for lesser included variants of the New York law.

One such bill is still under active review in California, another due for renewed consideration in the next session in Minnesota. Florida considered such a law in this year's session. In each of these cases, more than trivial support for legislation, none of it as yet enacted, came from education committee members and members on the legislative floor only after acceptance of provisions significantly less harsh than New York law or H.R. 1662.

Legislature study of varied problems of developing and producing tests led to reduction in the demand that every test form immediately be disclosed, so long as a substantial right to disclosed testing was preserved for those desiring that service.

Careful analysis of the need for research and technical information about tests led legislators to modify, eliminate or focus the "dragnet" style research filing requirements of the New York law.

Oregon, a State with an effective consumerist base, provided strong legislative support, still short of that needed for passage, for the 190 amended New York law. But, none of the many States evidencing strong interests in testing law has ignored the experience of the past 2 years.

None has sought to limit prematurely the self-regulating effort evident in most parts of the national testing community. While every State legislative proposal given serious consideration in 1981 has reflected the progress that has been made in the last year or two, H.R. 1662 has remained materially unchanged from its H.R. 4949 progenitor, drafted in 1979 after the New York law was enacted and before its effect produced radical changes in the procedures of many testing programs.

State legislators have learned that when it may be possible, indeed useful, in testing for admission to one career or profession may not be so for another. They have come in some measure to accept the differences in the character of tests and of their use in admission processes.

The Congress should avail itself of records, memoranda, and other products of hundreds of hours of State hearings that have taken place in the year since last we spoke of these matters in this hearing room.

I respectfully request that if the Congress is to consider a national testing law that it not disregard the major changes, both voluntary and coerced, that are so rapidly taking place in the admissions testing environment. As long as no State adopts a conflicting law, rather than a lesser included law, New York will provide a national standard admission testing law.

Steadily, many testing organizations voluntarily and by coercion are adopting national test disclosure policies. The forces that are getting them there should be allowed a bit more time to take effect.

If, in a year or two, the progress hitherto made toward voluntary openness in admission testing has not been maintained, or if in that time conflicting State laws and regulations have actually been enacted, the Congress might then consider a preemptive Federal testing act.

We will be pleased, at that time, to assist in the effort.

Thank you very much.

[The prepared statement of Bruce Zimmer follows:]

PREPARED STATEMENT OF BRUCE ZIMMER, EXECUTIVE DIRECTOR, LAW SCHOOL ADMISSION COUNCIL, WASHINGTON, D.C.

The Law School Admission Council is a nonprofit educational association whose members are the 171 law schools approved by the American Bar Association. Directly and through our operating corporation, Law School Admission Services, we provide a wide array of services and publications related to law school admission. Our most administratively and operationally complex program is carried out by our own corporation. It is the Law School Data Assembly Service, the national system of gathering and distributing academic and biographical information for the law school admission process. By virtue of our substantial effort to render diverse academic records readily comparable, this system has been instrumental in encouraging the consideration and evaluation of applicant academic records in the admission process. Each year, information concerning 100,000 people from 3,000 colleges applying to over 250 law schools is collected, processed and communicated through this system. Our many other varied activities aimed at the support and improvement of the law school admission process are described in Appendix B. The Law School Admission Test is another valuable component of the national law school admission program. We contract with the American College Testing Program and Educational Testing Service for the development of the Law School Admission Test. The test is administered under the auspices of our operating corporation, Law School Admission Services. On behalf of the Council I wish to thank this Committee for the opportunity to present our views and to respond to your questions.

Proposed testing legislation has impact upon a large number and wide variety of test sponsors, testing agencies, and test score users, as well as upon test-takers. No one can speak for all members of any of these categories. We speak only for ourselves and for the test-using law schools that are our members. No other test sponsor or test agency speaks for us. The Law School Admission Council does not support added legislation to regulate educational testing agencies or test use. We are not persuaded that there remain problems which demand legislative remedy, or that the asserted benefits of such legislation would outweigh the detriments of additional regulation of an educational community already burdened with extensive regulation. Further, the impact of voluntary and responsive actions of testing organizations is steadily becoming more apparent to our consuming and evaluating publics -- and especially to our critics. The recent record of test programs like ours may or may not document a need for some forms of regulation of educational testing. We welcome study of this question. We support the continued oversight of Congress in this matter. But we do not think a case for Federal admission test legislation has as yet been made.

The idea of federal testing legislation is not in itself unthinkable. Under certain potential circumstances a national law could preempt conflicting, confusing and oppressive state enactments. One appropriate and coherent scheme of regulation would be preferable to many uncoordinated and burdensome schemes. Indeed, our own research has led us to conclude that a sound model preemptive act could readily be developed. But the circumstances we face today suggest that enactment of a federal test law in the near future would be premature and probably unnecessary.

The national debate on testing for higher education admission has come a very long way in the four or five years that testing bills have been receiving serious attention from state legislatures and from members of Congress and their staffs. The environment in which testing programs operate has changed dramatically since July 1979 when New York enacted a test disclosure law to become effective at the start of 1980. At different rates, many testing organizations have made major -- indeed what two or three years ago would have been thought revolutionary -- commitments to openness in their testing programs. The Law School Admission Council has offered test disclosure for all of our tests throughout the United States and throughout the world, other than some very limited special administrations, since the New York law took effect. The New York test law did not permit us to continue our historic method of making scores comparable from test-to-test, and still disclose every form. Our decision to adopt universal test disclosure thus triggered a significantly revised structure for our test. As a result of that change we are planning to introduce a revised LSAT in June 1982. It will be comprised of four scored reading and reasoning sections, two unscored pretest or equating sections and an unscored writing exercise for individual law school evaluation. A ten to fifty score scale will replace the current 200 to 800 scale (See Appendix A for description of the 1982 LSAT revision) We are presently contracting to develop an adequate inventory of this edition of the LSAT. Having developed such a supply, we plan to address ourselves to emerging methods and technologies in evaluation in order to provide to law school applicants in the late 'eighties and early 'nineties the best academic measurement that can be provided at that time.

Test disclosure is a service our test-takers desire, and for which they are willing to pay. Because consumer demand for disclosed tests had become so substantial (about sixty percent of our test-takers) by the second year of test disclosure, and because of an enhanced interest on our part in providing an extra increment of quality assurance for our test, the Law School Admission Council Board of Trustees elected to make test and answer record disclosure an automatic part of the reporting procedures for test takers, beginning with the 1981-82 test year. This is a culmination of basic commitments on our part. We are rebuilding our services to accommodate our consumers' interest in openness in testing.

Two years ago the State of New York took a decisive position on test disclosure. It insisted on immediate action on full test disclosure. A sensible scheme of phased applicability would have been more fruitful. By giving four or five months notice for radical change to organizations that had hitherto operated on much longer test development lead times, New York disrupted the orderly conduct of testing programs like ours. Its actions caused increases in applicant fees that probably would not have been undertaken on the initiative of testing organizations alone. The 1979 New York law produced substantial technical problems and some areas of unintended encumbrance and intrusion. Several of these problem areas were modified in 1980 in a cooperative effort in which our organization took an active part. Final analysis of the impact of disclosed testing on test results must await research that can be performed only after several years of disclosed testing.

In no small measure the New York Act has worked reasonably because of the prudent regulations and sensible enforcement procedures of the New York Education Commissioner and his staff. Indeed, as the various test programs adjust to disclosed testing there is a greater risk that ill-advised regulations and enforcement procedures will emanate from the state educational bureaucracies than there is a probability that substantive provisions of a New York-type law would, themselves, prove intolerable if enacted in another state.

There is little question that -- to date -- New York has effectively preempted the national testing issue. It has acted, in effect, to set a national standard for many test programs. As Pennsylvania has done in the area of agricultural product standards and as New York itself has done in the insurance law field, it has established a *de facto* national standard. New York has done this by providing a powerful impetus to increased test disclosure nationwide. That impetus has come by virtue of the need and desire of test sponsors to maintain testing opportunities for New York students and schools -- as well as from the momentum of public opinion arising as test disclosure increased scrutiny of the underlying soundness of admission testing. Testing organizations like ours have discovered that, on balance, test disclosure promoted public confidence in the integrity of the test instrument and the testing process. It became symbolic of our public accountability. Indeed, openness has helped us in our quality control efforts.

As other states have considered test legislation in the last two years, nearly all have deferred to the New York law and experience as reference points. Most have concluded that the considerable progress toward national test disclosure made by programs like ours -- steadily joined by a growing list of major test sponsors -- had obviated the need for state-by-state regulation. In many instances, test sponsors have accelerated their commitment to openness in order to show good faith and thus eliminate a need for potentially burdensome regulation. During this past year's sessions there was support in several state legislatures for lesser included variants of the New York law. One such bill is still under active review in California, another due for renewed consideration in the next session in Minnesota. Florida considered such a law in this year's session. In each of these cases, more than trivial support for legislation (none of it as yet enacted) came from education committee members and members on the legislative floor only after acceptance of provisions significantly less harsh than New York law or H.R. 1662. Legislative study of varied problems of developing and producing tests led to reduction in the demand that every test form immediately be disclosed -- so long as a substantial right to disclosed testing was preserved for those desiring that service. Careful analysis of the need for research and technical information about tests led legislators to modify, eliminate or focus the "dragnet" style research filing requirements of the New York law. Oregon, a state with an effective consumerist base, provided strong legislative support -- still short of that needed for passage -- for the 1980 amended New York law. But, none of the many states evidencing strong interest in testing law has ignored the experience of the past two years. None has sought to limit prematurely the self regulating effort evident in most parts of the national testing community. While every state legislative proposal given serious consideration in 1981 has reflected the progress that has been made in the last year or two, H.R. 1662 has remained materially unchanged from its H.R. 4949 progenitor, drafted in 1979 after the New York law was enacted and before its effect produced radical changes in the procedures of many testing programs. State legislators have learned that what may be possible, indeed useful, in testing for admission to one career or profession may not be so for another. They have come, in some measure, to accept the differences in the character of tests and of their use in admission processes. The Congress should avail itself of records, memoranda and other products of hundreds of hours of state hearings that have taken place in the year since last we spoke of these matters in this hearing room.

I respectfully request that if the Congress is to consider a national testing law that it not disregard the major changes, both voluntary and coerced, that are so rapidly taking place in the admissions testing environment. As long as no state adopts a conflicting law, rather than a lesser included law, New York will provide a national standard admission testing law. Steadily, many testing organizations voluntarily and by coercion are adopting national test disclosure policies. The forces that are getting them there should be allowed a bit more time to take effect.

If, in a year or two, the progress hitherto made toward voluntary openness in admission testing has not been maintained, or if in that time conflicting state laws and regulations have actually been enacted, the Congress might then consider a preemptive federal testing act. We will be pleased, at that time, to assist in the effort.

APPENDIX A

When the New York Standardized Testing Act was passed in 1979, the sponsors intended to preserve the testing organizations' ability to make scores comparable from test form to test form and from test administration to test administration. But their inquiry into the process of equating, as the above described process is called, was limited to one or two of the largest volume test programs. In equating, as in other areas of substance, admission tests are diverse in content, structure and goal. Test programs in law, graduate school, and business admissions were overlooked by the test sponsors' broad legislative brush strokes. Thus, the New York test statute, by virtue of its requirement that each test form used be published, required us to abandon a method of test production that had been effective for our test and to commence a crash research program to find another equating method. Until January 1, 1980, new editions of the LSAT were placed on the score scale by administering to comparable populations the new edition and, at the same time, editions already on the scale. In other words, whole editions of the LSAT were "spiraled" with other editions of the test at the same administration. Editions of the test were typically used at two or three different administrations. As a practical matter this production method was precluded by the statutory requirement that each test form used be published after every administration. The Law School Admission Council had at its disposal a very limited number of usable test forms which were placed on scale before The New York Test Law was enacted. These tests have been used to bridge the time gap until a new equating method can be developed and implemented. The introduction of a new method of equating required changes in the LSAT itself, as well as in the statistical procedures used to assure that scores achieved on one edition of the LSAT mean essentially the same thing as those achieved on another. Development of a revised Law School Admission Test, adapted to a test disclosure environment and encompassing other desired improvements is under way. The revised test, to be introduced in June 1982 is described in preliminary fashion in the materials that follow as part of this appendix. Please note that the July 1981 Law School Admission Bulletin will soon be supplanted by a more current revision. The material concerning Question Type III, in particular, will be subject to revision as the refinement of the test continues through its final developmental stages.

Route To: ____________
____
____

LAW SCHOOL ADMISSION BULLETIN

November 1980

Development of New LSAT Proceeds

Development work is under way on the most significantly different Law School Admission Test in the thirty-two year history of that test. The Law School Admission Council Board of Trustees approved this major effort at its Spring 1980 meeting. Approval followed a series of recommendations made by the Council's Test Development and Research Committee after thorough study. As much as two million dollars is to be expended on the development and introduction of fifteen new forms of the LSAT. The LSAC Board reported on the decision and its implementation during the educational component of the LSAC Annual Meeting in late May, 1980. The new test is expected to be given for the first time in the summer of 1982. That test is expected to:

¶Consist of six thirty-five minute sections, two of which will be equating or pretest sections that will not contribute to candidate scores. The four scored sections will include four different types of questions, only one of which is currently used.

¶Measure abilities and skills related to problem identification, analysis, logic, reasoning and reading.

¶Also contain, subject to modification based on future research, an unscored writing sample or exercise. A copy of the completed exercise would be transmitted to each law school to which a candidate directed an LSDAS/LSAT report.

¶Be equated, that is have scores statistically made to "mean the same thing," only to other new forms of the test and not to the present LSAT.

¶Be placed on a score scale different from the 200-800 scale currently employed.

Even though scores on the new LSAT will not be directly comparable to scores on the present test, the design specifications have been set to cause the test to have psychometric qualities which will achieve at least the levels of performance of the present test. This means that the new LSAT will be designed to distinguish among law school applicants at various ability and talent levels as well as the present test does.

Questions for inclusion on the new test are now being developed at Educational Testing Service under contract with Law School Admission Services. General direction and policy decisions related to the project are vested in the LSAC Test Development and Research Committee. Day-to-day management of specific components of the work is being carried out by relevant LSAC/LSAS professional staff. Questions for the new test are being pretested in current administrations of the LSAT. The initial phases of the project are proceeding according to schedule.

Many of the contemplated changes have been under consideration by the Test Development and Research Committee for the past few years. These changes are of a kind that will permit the retention of the principal design considerations of the past -- strong validity, high reliability and difficulty, and good differentiation among persons of high levels of ability and training.

Writing Ability Score Eliminated

The current Writing Ability score will be eliminated when the new version of the LSAT becomes operational. There are a number of reasons for this change. First, the WA score adds little or nothing to prediction of academic performance for the great majority of law schools. Second, "Writing Ability" is a misleading title, given the type of question on which the score is based. Finally, no type of question testing writing skills in a multiple-choice format has been judged particularly desirable for inclusion in the new version of the LSAT.

LSAC has approved, subject to future research considerations, an unscored writing sample or exercise for the new version of the LSAT. Copies of a candidate's written work on this exercise would be forwarded to law schools that receive his or her LSAT score. Such an exercise would be intended to stress to candidates the importance of writing in legal education. It should enable a law school to make an informed judgment concerning the basic writing skills of a particular applicant.

Increased Law Faculty Participation

In order to provide a substantial opportunity for law faculty reaction to specific questions proposed for the new test, LSAC President Walter B. Raushenbush is appointing a Test Question Review Committee. That Committee, a diverse group of law professors, is to begin its work in January, 1981. It will review, analyse and critique prospective LSAT questions. The goal of the review project is to assure that every question on the new LSAT has been screened by experienced law professors. The committee will review questions for clarity, fairness and other relevent tests of appropriateness.

Background for Changes: New York Law

In 1980, legislation regulating post-secondary admission testing programs became effective in New York. Similar legislation is being considered in twenty states and in Congress. One result of the legislation is to require LSAC/LSAS to make changes in the way in which the LSAT is constructed and administered. The New York Testing Act affects many aspects of the LSAT that have been under discussion since 1977 by the Test Development and Research Committee of LSAC as well as staff and consulting professional advisers.

Equating, Test Production and Statutory Compliance

Because it requires the disclosure of each form of the LSAT (other than those used at limited administrations) the New York Testing Act makes it necessary to develop a new method of equating the test. Equating is the statistical process by which scores attained on different forms of a given test are adjusted to take into account slight differences in difficulty. By equating, differences in scores which might result from a person having taken one form of the LSAT rather than another are minimized. In the past, a new form of the LSAT was equated by administering the new form along with two previously equated forms of the test. Approximately one third of the group taking the test used the new form and one third each of the previously equated forms. Difficulty levels and scaled scores for the new test could then be determined by comparing the performances of the three randomly selected groups of test takers. This method of equating a new form of the LSAT is generally known as spiraling.

As you can see, the New York Law precludes the use of the spiraling equating method. Because the law requires the disclosure of every test form immediately following the use of that form, there can be no previously equated and undisclosed forms to spiral with new forms of the LSAT, once the present supply is exhausted. A new method of equating needs to be found if the LSAT program is to continue in the same manner as in the past. Of equal concern is the need to produce many more new forms of the LSAT than in the past. Until disclosure of each form of the test used was required, a form of the test was usually administered three or four times before being retired. With disclosure of every form of the test administered no form could be reused without risking compromise of the accuracy and fairness of the testing process. Production of the new LSAT forms will have to exceed by a considerable amount the number of forms required in the past. In turn, this leads to concern for the quality and quantity of questions.

The needs to find a new equating method, to maintain the quality level of the LSAT and to produce more editions of the LSAT every year dictate changes in both the types of questions being used on the LSAT and in their presentation. These same changes will also result in an increase in flexibility in the way the LSAT is administered and provide an additional method to deal with the concern for the test taker who may be tempted to copy from another. Applicants sitting next to each other during a test administration will be less likely to be tempted to copy from those seated beside them if they know that they are working on different sections of the test at the same time, as they may well be.

Pretesting Methods

As just noted, sections of equal length will make possible "scrambling," or re-ordering, of test sections, on the same form and among various forms of the test. This process will permit pretesting of different sections of the test at different places in the test.

All LSAT questions are pretested before being used operationally on a form of the test. Among other reasons, questions are pretested to establish their level of difficulty and to identify any that appear to operate poorly or differently than anticipated. Answers to questions being pretested do not contribute to candidates' LSAT scores and, consequently, the New York Testing Act does not require that these questions be disclosed. Reviews of recent forms of the LSAT produced information that pretest statistics for some questions differed from the statistics demonstrated when the same questions were included as operational items on the test. Some explanations of this phenomenon are plausible. Until this year, pretest sections appeared in the same place on all forms of the LSAT. This fact was known by many candidates. Candidates may have been tempted to "loaf" or "let up" on pretest sections. If they did, their performance on pretest questions did not accurately reflect the way in which they would approach operational questions. Since the objective of the pretesting process is to gather accurate information about the performance of questions being considered for use in the LSAT, the practice of consistently placing pretest questions in the same positions on each test might have detracted from the reliability of the information produced.

Reducing Potential for Coachability

The effects of coaching — the use of artificial or short-term learning to improve test performance -- have also been considered in the design of the new LSAT. Coaching success may be related to question or item type format, particularly to the "fixed format" nature of the question involved. Questions being considered for the new version of the LSAT have been examined for potential coachability. Only those types of questions for which success in coaching has been judged to be low will be selected.

The Score Scale

The 200-800 score scale of the LSAT was established in 1948. It was based on the performance of candidates who took a test far different from the one now being administered. Over time, new versions of the LSAT, containing different types of questions, have been introduced gradually and equated to previous versions. Despite these changes, the 200-800 score scale was preserved.

Because of the major change in questions employed on the new LSAT, that version will not test the same qualities that were tested by older versions in exactly the same ways. Consequently, even if the 200-800 scale were to be retained, scores on the old and new versions would not likely mean the same thing. While there is no change in the criterion for the test -- or the intended use of the test, groups of applicants might be ordered slightly differently by the new test than they would have been on the current LSAT. In addition, the particular 200-800 score scale employed in the LSAT program has been a subject of concern to LSAC for some time. This three-digit score scale can create an impression of precision that is not warranted.

The problems described above--masking the non-comparability of old and new versions of the LSAT by using the same score scale, and creating a misleading appearance of precision with the current 200-800 score scale--can most effectively be resolved by establishing an entirely different score scale for the new version of the LSAT being developed. The LSAC is actively discussing alternative score scales. No final decisions have yet been made concerning the score scale that will be adopted for the new version of the test. It is very likely that scores earned by candidates taking the new version of the test will look quite different from scores that have been reported on the 200-800 scale in the past. Scores will probably be reported by use of fewer digits, fewer score intervals, and a slightly smaller range than is the case with the current LSAT.

Further Reports

As work on a new LSAT progresses, periodic reports will be made in this publication, as well as at various meetings and programs of LSAC.
GLD/WBR
TOW/BIZ

Route To: ____________

LAW SCHOOL ADMISSION BULLETIN

January 1981

Board Decisions Advance Program For New LSAT

In the November issue of this publication a comprehensive report on the development of a new LSAT was offered. At its December meeting, the LSAC Board decided significant issues related to the implementation of that project.

A new score scale has been selected for the test currently being developed. The new LSAT will be scored on a 10 to 50 scale. The mean score will be 30, the standard deviation will be 8. Thus about two-thirds of scores will occur in the 22 to 38 range, in contrast to about the 450-650 range on the present test. The scale will utilise 2 1/2 standard deviations from the mean, rather than 3 as on the present test. This means that about one percent of the top and bottom performances will be compressed downward or upward respectively.

Readers of the Bulletin will recall that a scale other than 200-800 was selected for the new LSAT because that instrument will not test the same qualities that were tested by older versions of the LSAT in exactly the same ways. Consequently, even if the 200-800 scale had been retained, scores on the old and new versions would not likely mean the same thing. While there is no change in the criterion for the test -- or the intended use of the test, groups of applicants might be ordered slightly differently by the new test than they would have been on the current LSAT. In addition, the three digit score scale employed in the LSAT program has been a subject of concern to LSAC for some time because the three digits can create an impression of precision that is not warranted.

In order to preserve current expectations of applicants and law schools, both old and new LSAT scores will be reported for a five-year period after the introduction of the new LSAT. And finally, the Board ratified the use of six 35 minute sections, two of which will be variable content pretest or equating sections. A writing sample section is to be included, however its length has not yet been determined.

The calculation of an LSAT College Mean (LCM) has for some time been controversial -- both as to reliability and utility. The Board has tentatively decided to abandon LCM calculation with the introduction of a new test. A fuller discussion of the LCM issue will take place at the 1981 Annual Meeting.

(APPENDIX B)

LAW SCHOOL ADMISSION COUNCIL

The Law School Admission Council (LSAC) is a non-profit educational association of the 171 ABA-approved law schools organized to facilitate the admissions process. According to its charter of incorporation, the Council's purposes are "to construct, administer, and report scores for tests for admission to law school; to conduct educational research; and to provide services to law schools and the educational community." As the number of applicants has increased, the law school admission process has become more competitive and complex. The range of Council services and activities has expanded to serve and inform those involved in that process--law schools, applicants and undergraduate institutions. Services, programs and publications are provided by the Council to assist each of these groups.

History. In 1947, representatives of a group of law schools met with College Entrance Examination Board staff to discuss the possibility of developing a common test for use in admitting students to law school. There was considerable interest in proceeding with the project, and efforts to develop such a test were undertaken by the Educational Testing Service. The first Law School Admission Test (LSAT) was administered in February, 1948. Nineteen law schools used the test that year. Those schools became associated as the Law School Admission Test Council, with a Policy Committee to set operational guidelines for the new testing program. In later years, additional committees were established to oversee aspects of the program. The Policy Committee evolved into the Executive Committee and then the Board of Trustees. In 1968 the Law School Admission Test Council was incorporated and in 1973 the corporate name was changed to "Law School Admission Council" to reflect the broader activities of the organization. In 1978 the Council opened a national office in Washington, D.C. to conduct public affairs and government relations work and to provide support for numerous Council activities. In 1979, the Council incorporated a wholly-owned not-for-profit subsidiary called Law School Admission Services (LSAS), governed by a nine person Board of Directors appointed by the LSAC Board. LSAS, operating as of July 1980 in Newtown, Pennsylvania, has full responsibility for operation of the Law School Data Assembly Service, and all other Council services other than the development and administration of the Law School Admission Test. ETS remains, at present, contractor for these aspects of the LSAT.

Governance. The Council conducts formal business at an annual meeting, held in recent years in late May. Member schools have one vote apiece, cast by a Council representative selected by the dean or faculty.

The ongoing governance of the affairs of the Council is conducted by its Board of Trustees. The Board normally meets three times a year (in early December, in late April or early May, and at the Annual Meeting) and has fifteen members: the President; the Immediate Past President or the President-Elect in alternate years; the chairpersons of the five standing committees; six members elected by the Council (three each year) for two years. The term of the President is two years; the President-Elect becomes the President on completion of one's service as President-Elect. The Secretary is elected for a one-year term by the Board of Trustees at its meeting immediately following the Annual Meeting.

Staff. The Executive Director of the Council operates an office in Washington, D.C. which provides support for a variety of Council activities. A substantial LSAS staff is maintained in Newtown, Pennsylvania. It operates under the direction of the President of that organization.

Finances: Virtually all of the Council's income is derived from fees paid by those who register for the basic admission services (LSAT and LSDAS) and those who purchase the Prelaw Handbook. The Council's operating budget for 1980-81 is about six and one half million dollars.

Committees. The Council has five standing committees (Finance; Legal Affairs; Programs, Education, and Prelaw; Services; and Test Development and Research) The chairpersons and members of these committees are appointed by the President and are responsible to the Board of Trustees. Policy recommendations and appropriation requests from committees are generally referred to the Board for final action, although matters requiring an immediate decision can be handled in consultation with the President. Standing committees meet at least twice a year on a regular basis.

Organizational Affiliations. The Council is a founding and continuing sponsor of the Council on Legal Education Opportunity and has working relationships with many organizations active in legal education.

Annual Meeting. The Annual Meeting, which has been held in recent years during the last week of May, includes workshops, seminars, and a business meeting. Workshops and seminars cover topics of current concern to those responsible for admissions--e.g., affirmative action, statistical concepts, legal and legislative developments and use of non-quantifiable factors. At the business session, reports are considered and officers and new trustees elected.

Other Meetings and Workshops. The Council sponsors several educational programs during the year, including:

Regional Workshops are one day programs designed for law school faculty members who serve on admissions committees and administrators. The regional workshops are scheduled during the fall and are designed to explore major issues in depth.

The Summer Workshop is a 3½ day program in late June or early July, designed for faculty members and senior law school administrators with responsibility for admissions. Significant subjects are considered in large general sessions and small discussion groups.

Prelaw adviser workshops are one day programs for those in undergraduate institutions who advise students contemplating admission to law school. The workshops are designed to acquaint advisers with Council services and with various aspects of the admissions process.

Services. The LSAT is a half-day multiple choice test designed to measure abilities important in the study of law. It provides two scores: an LSAT score and a writing ability (WA) score. The LSAT portion measures the ability to understand and reason with a variety of verbal and quantitative materials. The writing ability portion measures the use of standard English. The test is intended to supplement the undergraduate record and other information about the student in the assessment of potential for law school work. It covers a broad range of disciplines, measures skills usually acquired over a long period of time, and gives no advantage to students with particular specializations.

The Law School Data Assembly Service (LSDAS) centralizes the collection and analysis of college transcripts required for law school admission decisions and combines a summary of these transcripts with LSAT scores and background information (e.g., address and date of birth) to produce an LSDAS Report for each candidate. The LSDAS Report contains a summary of all the data collected (with grades and measures of credit reported in a standard form of notation), is attached to copies of the candidate's transcript(s), and is forwarded to law schools designated by the candidate.

The Law School Candidate Referral Service (LSCRS) provides law schools approved by the ABA an opportunity to have the LSDAS file searched to identify candidates who have given their permission to be in LSCRS and who have characteristics specified by the schools participating in the service. It thus provides students an opportunity to be made aware of educational and scholarship possibilities that they might not otherwise have considered.

The Council also sponsors services for internal use by law schools in managing their admissions operation--e.g., Extended Services Reports (a series of status and statistical reports produced periodically from the LSDAS system) and the Validity Study Service (which provides for continuing evaluation of the effectiveness of LSAT scores and undergraduate grade point averages in predicting academic performance in law school).

Publications. Several Council publications--e.g., the Guide to Undergraduate Colleges (which provides statistical data and information about the grading systems of undergraduate institutions), the Operations Reference Book (which contains operational information about Council services), the Validity Study Service binder (which includes reports of an individual law school's validity studies and general background material), the Directory of Prelaw Advisors (which lists the names and addresses of advisers at undergraduate institutions), and a widely distributed newsletter published by both LSAC and LSAS, Law School Admission Bulletin. Other publications, listed below, are more general in nature.

The annual publication of registration materials is intended for candidates who take the LSAT and use the LSDAS. It explains these and other Council services, provides complete registration information and forms, and discusses the meaning of scores and contains some sample test questions. Copies are available without charge. In recent years, the materials have been expanded to include an actual LSAT form with an answer key, information on preparing for the test, general test-taking strategies, and materials on how to answer LSAT questions. It also includes a discussion of the different types of test questions and an analysis of the correct answers for some sample questions.

Reports of LSAC Sponsored Research is a three volume set of research reports completed before December 1977. In addition, a green looseleaf binder entitled "LSAC Research" contains a brief overview of all Council sponsored research, and a summary of each completed project. As new projects are completed, a summary and full report will be issued to update the binders. Two sets of binders and bound volumes have been sent to each law school--for the law library and one for the admissions office.

The Legal Affairs Manual, published jointly by the Council and the Association of American Law Schools (AALS), contains a checklist of federal statutes and regulations which affect the operations of law schools, memoranda analyzing the impact of particular statutes and regulations on law schools and suggesting ways in which schools might respond, the pertinent federal statutes and regulations, and a litigation bulletin summarizing state and federal cases of particular interest to legal education. The Manual and its supplements are distributed without cost to all Council member schools; the subscription rate for others is $50 for the original binder and an annual $25 fee for the up-date service.

The Annual Council Report contains financial reports and a variety of other materials reflecting Council activities for the year. It is distributed to all member law schools.

The Prelaw Handbook is a publication of the Council and the AALS. It contains general information on the legal profession and law schools for those interested in the study of law, a description of the program of each American Bar Association approved law school, and the admissions profile for the most recently admitted class for the great majority of these schools. It may be purchased at most college bookstores or by mail from Law School Admission Services.

Law as a Career/Prelaw Advice for Minority Students. The first of these pamphlets is written for high school seniors and college underclassmen, while the second is directed to minority students. Both are available from LSAS in quantities of 100 for $6.00.

The "Prelaw Adviser Issue" of the Law School Admission Bulletin is published quarterly to provide prelaw advisers with information about Council services and general admission matters.

The Prelaw Advisor Kit is a portfolio for advisers, containing materials explaining in detail Council services and policies. Also included are copies of several Council publications. A Kit is sent annually without charge to each prelaw adviser; additional copies are available for $5.00 each from LSAS.

Route To: ______

LAW SCHOOL ADMISSION

BULLETIN

July 1981

LAW SCHOOL ADMISSION TEST
Preliminary Descriptive Material For Question Types For 1982 Revision

In General

The LSAT is designed to measure certain mental abilities important in the study of law and, thus, to aid law schools in assessing the academic promise of their applicants. The test covers a broad range of academic disciplines and is intended to give no advantage to candidates from a particular academic background. [illegible] questions yielding the LSAT score [illegible] designed to measure the ability to read, understand, and reason.

In the Spring of 1980, final approval was given by the Board of Trustees to the Law School Admission Council for the introduction of a revised version of the Law School Admission Test. This "new" version of the LSAT is scheduled to be administered for the first time in June, 1982.

Even though scores on the new LSAT will not be directly comparable to scores on the present test, the design specifications have been set to cause the test to have psychometric qualities which will achieve at least the levels of performance of the present test. This means that the new LSAT will be designed to distinguish among law school applicants at various ability and talent levels as well as the present test does.

The Questions

The "new" LSAT will consist of six thirty-five minute sections, two of which will be equating or pretest sections that will not contribute to an examinee's score. The four scored sections will include four different types of questions. It is anticipated that these four sections will consist of between 110-120 questions.

Also planned is a 20-30 minute writing sample or exercise. This writing sample will not be scored. A copy of the examinee's written work will be sent to each law school to which an LSAT/LSDAS Report is sent for the examinee.

The Score Scale

The 200-800 score scale of the LSAT was established in 1948. It was based on the performance of candidates who took a test far different from the one now being administered. Over time, new versions of the LSAT, containing different types of questions, have been introduced gradually and equated to previous versions. Despite these changes, the 200-800 score scale was preserved.

Because of the major change in questions employed on the new LSAT, this version will not test the same qualities that were tested by older versions in exactly the same ways. Consequently, even if the 200-800 scale were to be retained, scores on the old and new versions would not likely mean the same thing. While there is no change in the measurement objectives for the test, or the intended use of the test, groups of applicants might be ordered slightly differently by the

Copyright 1981 by Law School Admission Services, Inc. All rights reserved.

new test than they would have been on the current LSAT.

In addition, the particular 200-800 score scale employed in the LSAT program has been a subject of concern to LSAC/LSAS for some time. This three-digit score scale can create an impression of precision that is not warranted.

It has been determined that an entirely different score scale will deal most effectively with these and the many other considerations which must be taken into account when making a decision about the scale to be used on a test such as the LSAT. The scale decided upon will yield scores across the range 10-50. The lowest score will be 10, and 50 will be the highest score. It is expected that these extremes will be approximately two and one half standard deviations from the mean.

Preparation Materials

LSAC/LSAS recognizes that there is interest in and concern about this revision in the LSAT, even though it will not be administered until June, 1982. Of particular concern are the types of questions which will be included on the "new" test. The descriptions of the four question types which will comprise the "new" LSAT provided here are intended to meet these interests and concerns at this early state of introduction of the "new" test. Additional information about the test and its introduction will be made available during the 1981-82 school year.

On the following pages there are sample questions similar to those which will be used in the revised version of the LSAT, and discussions of the answers, including explanations of how the answers were determined. For each type of question there is first a general description and discussion of the objectives and techniques for that question type. Then follow the directions preceding that type of question and sample questions with discussions of the answers. The general level of difficulty of most of these questions is given to provide a guide to how examinees have performed on these questions.

As with the present LSAT, actual test forms will contain a number of trial questions, and all examinees do not necessarily receive exactly the same test book. Answers to trial questions will not be counted in the scoring.

The descriptive materials represent the general nature of the test. LSAC/LSAS have tried to present information here about the types of questions which will be used in the LSAT. But, circumstances may require that other question types be used. Thus the caution that, material not described in the discussions of question types may appear in the test.

The "General Test-Taking Suggestions" in the 1981-82 LSAS General Information Booklet will be appropriate to the revised version of the LSAT.

In the questions which follow the ideas, presentations and views are those of the original authors and are not intended to represent the opinions of Law School Admission Council, Law School Admission Services or their agents and contractors.

Please remember that the "new" LSAT is in the final stages of development and that some changes in the material presented here are expected.

Question Type I

The purpose of this question type is to measure the ability to read with understanding and insight. This type of question explores the examinee's ability to analyze a written passage from several perspectives: the ability to recognize both explicitly stated elements in the passage and assumptions underlying statements or arguments in the passage as well as the implications of those statements or arguments. Because the written passage upon which questions are based presents a sustained discussion of a particular topic, there is ample context for analysing a variety of relationships; for example, the function of a word in relation to a larger segment of the passage, the relationships among the various ideas in the passage, or the relationship of the author to his or her topic or to the audience.

There are six specific kinds of questions related to this type. These kinds focus on (1) the main idea or primary purpose of the passage; (2) information explicitly stated in the passage; (3) information or ideas implied or suggested by the author; (4) possible application of the author's ideas to other situations; (5) the author's logic, reasoning, or persuasive techniques; and (6) the tone of the passage or the author's attitude as it is revealed in the language used.

This question type contains passages of approximately 450 words, each passage providing the basis of answering six or more questions. The passages are drawn from a variety of sources and subject matter areas.

Suggested Approach

¶Since reading passages are drawn from many different disciplines and sources, you should not be discouraged by encountering material with which you are not familiar. It is important to remember that questions are to be answered on the basis of the information provided in the passage. You may, however, want to do last a passage that seems to you particularly difficult or unfamiliar.

¶There are different strategies for approaching the questions; you should decide which works most effectively for you. You might experiment with different strategies. Some different strategies are: reading the passage very closely and then proceeding to the questions; skimming the passage, reading quickly through the questions, and then rereading the passage closely; and reading the questions first, then reading the passage closely. You may find that different strategies work better for different kinds of passages; for example, it might be helpful with a difficult or unfamiliar passage to read through questions first.

¶Whatever strategy you choose, you should analyse the passage carefully before answering the questions. As with any kind of close and thoughtful reading, you should be sensitive to clues that will help you understand less explicit aspects of the passage. Try to separate main ideas from supporting ideas or evidence; try also to separate the author's own ideas or attitudes from information he or she is presenting in an objective manner. It is important to note transitions from one idea to the next and to examine the relationships among the different ideas or parts of the passage: For example, are they contrasting? Are they complementary? You should consider the points the author makes and the conclusions he or she draws and also how and why those points are made or conclusions drawn.

¶You may find it helpful to underline or mark key parts of the passage. For example, you might underline main ideas or important arguments, or you might circle transitional words (although, nevertheless, correspondingly, and the like) that will help you map the structure of the passage or you might note descriptive words that will help you identify the author's attitude toward a particular idea or person.

¶Always read all the answer choices before selecting the best answer.

¶Be careful not to pick an answer choice simply because it is a true statement; be careful also not to be misled by answer choices that are only partially true or only partially satisfy the problem posed in the question.

¶Answer the questions on the basis of the information provided in the passage and do not rely on outside knowledge. Your own views or opinions may sometimes conflict with the views expressed or the information provided in the passage; be sure that you work within the context provided by the passage. You should not expect to agree with everything you encounter in reading passages.

Sample Questions

Each passage in this group is followed by questions based on its content. After reading a passage, choose the best answer to each question and blacken the corresponding space on the answer sheet. Answer all questions following a passage on the basis of what is stated or implied in that passage. (Some or all of the passages for this test have been adapted from published material to provide the examinee with significant problems for analysis and evaluation. To make the passages suitable for testing purposes, the style, content, or point of view of the original may have been altered in some cases.)

There are two social conceptions which can be expressed only in terms of myth. One is the social contract, which presents an account of the origins of society. The other is the utopia, which presents an imaginative vision of the telos or end at which social life aims. The myth of the social contract and the myth of the utopia both begin in an analysis of the present, the society and concerns that confront the mythmaker, and they project this analysis in time or space. The contract projects it into the past, the utopia into the future or to some distant place. To Hobbes, a contemporary of the Puritan Revolution, the most important social principle was the maintenance of de facto power; hence he constructs a myth of the social contract that turns on society's surrender of that power. To Locke, a contemporary of the Whig Revolution, the most important social principle was the relation of de facto to legitimate or de jure authority; hence he constructs a myth of the social contract that turns on society's delegation of power. The value of the myth of the social contract as social theory depends on the depth and penetration of the social analysis that inspires it.

The social contract, although it is a genuine myth that disguises opinion as fact, is also usually regarded as an integral part of social theory. The utopia, on the other hand, although its origin is also social analysis, is thought of as fiction. The reason for the difference in the way the myths are perceived is that the emphasis in the contract myth falls on the present facts of society, facts that it is supposed to explain. And even to the area where historical evidence lies; and so the myth preserves at least the gesture of making assertions that can be definitely verified or refuted.

The utopia, however, cannot even pretend to make verifiable assertions because it is an overtly speculative myth; it is designed to contain or provide a vision of one's social ideas, not to be a theory connecting social facts together. There have been one or two attempts to take utopian contructions literally by trying to set them up as actual communities, but the histories of these communities make melancholy reading. The utopian writer looks at his or her own society first and tries to see what, for his or her purposes, its significant elements are. The utopia itself shows what a society would be like if those elements were allowed to exist in their ideal form. Plato looked at his society and saw its structure as a hierarchy of priests, warriors, artisans, and servants--much the same structure that inspired the cast system of India. The Republic shows what a society would be like in which such a hierarchy functioned on the principle of justice, that is, each person doing his or her own work.

1. The author believes that the myth of the social contract can be valuable as social theory when the writer of the myth

   (A) has studied the history of the myth of the social contract
   (B) has analyzed the differences between the conceptions of the social contract and the utopia
   (C) has made an insightful analysis of society
   (D) is concerned with the justification of power
   (E) is interested in social goals

This question refers specifically to a statement made by the author at the end of the first paragraph. The author says that the value of the myth of the social contract depends on the quality--the "depth and penetration"--of the analysis of society that forms a background to the particular myth. The answer to this question of moderate difficulty is, therefore, choice C. None of the other answer choices relates directly to the author's view of what makes the social contract valuable as social theory. While the author might agree that the activities described in choices A and B and the kinds of motivation suggested in choices D and E might affect the nature or value of a particular myth of the social contract, none of these four choices directly or adequately answers the question as it is posed.

2. With which of the following statements about an actual utopian community would the author be most likely to agree?

   (A) A utopian community is more equitable to all of its members than is a non-utopian community.
   (B) Life in a utopian community is not likely to be much more pleasant than life in present-day society.
   (C) In the future there will be no more utopian communities; they will be found only in books.
   (D) A community based on utopian ideals is more likely to survive than is one based on the concept of the social contract.
   (E) Although it is unlikely that a utopian community will flourish in the future, any attempt to found a utopian community should be encouraged.

In order to answer this rather difficult question, you must evaluate each of the answer choices in reference to the author's various statements about actual utopian communities. The statement that the history of actual utopian communities makes "melancholy reading" implies that the author believes that historical utopia communities have generally failed to achieve their goals. Choice A can be eliminated because there is no evidence in the passage that utopian communities have, in the author's view, achieved greater equity than ordinary communities. Choice B, however, is compatible with the idea that actual utopian communities do not accomplish their goals, and it is the best answer. Choice C, while at first appearing plausible given the author's lack of optimism with regard to the success of these communities, is not supported by any evidence in the passage; the author nowhere indicates or implies that utopian communities are unlikely to be established in the future. Choice D can be eliminated because there is no evidence in the passage to support it and because the author's description of the social contract in fact suggests that a society based on "present facts" or analysis of the past might be more viable than one based on the imaginative "fiction" of the utopia. Although choice E acknowledges the author's skepticism with regard to the utopian communities, it incorrectly attributes support of utopian communities to the author; a positive view of the worth of such communities is not borne out by anything in the passage.

3. According to the passage, Plato is an example of a mythmaker who

   (A) contructed a utopian society based on elements of his own society
   (B) contructed a utopian society as a way of criticizing his own society
   (C) contructed a utopian society set in a period before recorded history
   (D) used the myth of the utopia to convey the concept of the social contract
   (E) used the myth of the utopia to make verifiable assertions about history

This question refers to the last two sentences of the last paragraph of the passage. There the author describes the structure of Plato's society and indicates that The Republic projects that society into one operated "on the principle of justice" - a utopia. Once the links between Plato and The Republic and a society operated "on the principle of justice" and "a utopian society" have been made, choice A can be seen to be the best answer. While each of the other choices might be accurate statements there is no information in the passage to support their accuracy and thus none of them is preferred over A. This question is of moderate difficulty.

4. In the passage the author is primarily concerned with

   (A) analyzing a process
   (B) developing an analogy
   (C) explaining differences
   (D) settling arguments
   (E) refuting opponents

This relatively easy question requires a choice among very brief summaries of the purpose of the passage. Differences in the myths are directly referred to in the passage and examples and evidence of contrast are the constant focus of writing. All of the choices other than C refer to activities in which the author expresses no interest nor is in any way involved in the passage.

5. According to the passage, which of the following best describes the difference between the way in which the myth of the social contract and the myth of the utopia are perceived?

   (A) The myth of the social contract is often regarded as an accurate explanation of present society, whereas the myth of the utopia is often regarded as a fiction involving elements of present society.
   (B) The myth of the social contract is often believed to be an accurate account of society's past, whereas the myth of the utopia is usually regarded as a probable account of society's future.
   (C) The myth of the social contract is usually regarded as a less imaginative treatment of the themes that are the basis of the myth of the utopia.
   (D) The myth of the social contract is usually thought of as a political statement, whereas the myth of the utopia is usually thought of as a sentimental treatment of social goals.
   (E) The myth of the social contract is usually thought of as

insightful social analysis, whereas the myth of the utopia is usually thought of as a psychological analysis.

This question requires the review of answer choices which purport to sythesize and integrate views found in the passage. Choice E can be eliminated by virtue of the reference to psychological analysis as the way in which utopia is usually thought of. No direct support for this conclusion is found in the passage. Both the references to political statement and sentimental treatment in choice D are very difficult to support based on anything directly stated or suggested by the author in the passage and render the choice weak. Choice C suggests a comparison based on how imaginative the themes of the myth are perceived to be. This contrast is no where directly stated and for C to be the preferred answer would require you to infer an evaluation of social contract as to the level of imagination shown by it since the myths are not directly compared by the author in this way. It too becomes a weak choice. Choice B is not preferred because it states that social contract is an accurate account of the past while the passage declares that it emphasizes the present and because it declares that utopia is perceived as being probable though the passage indicates it to be an ideal with no suggestion of its probability. The best choice for this difficult question is A since it directly reflects the first part of the second paragraph of the passage.

6. The primary purpose of the passage is to

(A) offer some reasons for the continued popularity of the myth of the social contract and the myth of the utopia
(B) show that political events have influenced the writers who have written about the concepts of the social contract more than they have influenced the writers who have written about the concept of the utopia
(C) compare the origin of the myth of the social contract with the origin of the myth of the utopia on the basis of the interrelationship between political and economic factors
(D) explain why there is a difference between the way in which the myth of the social contract has been perceived and the way in which the myth of the utopia has been perceived in the past
(E) contrast the myth of the social contract and the myth of the utopia on the basis of the way in which each type of myth is related to social analysis and the way in which each type of myth is perceived.

Choice A in this question refers to the continued popularity of the myths which is no where addressed by the author in the passage. While political events are made in the passage to establish reference points for Locke and Hobbes, no contrast between these writers and those dealing with utopia is ever offered by the author. Thus choice B is a weak one. The comparison suggested in Choice C is also not made by the passage which states that the origin of both myths is in the present but does not focus on political, economic or other specific factors in society in its references to the origin of the myths. Choice C is not preferred. Choice D is not preferred because it deals only with the perception aspect of the passage while choice E integrates both perception of the myths and their relationship to social analysis and is the best answer to this moderately difficult question.

Question Type II

Type II questions focus on the ability to understand a structure of relationships and to draw conclusions about that structure. The examinee is asked to understand the conditions used to establish the structure of the relationship and to deduce new information from them. Each group of Type II questions consists of (1) a set of some three to seven related conditions (and sometimes other explanatory material) describing a structure of relationships, and (2) three or more questions that test understanding of the implications of that structure. Although each question in a group is based on the same set of conditions, the questions are independent of one another; answering one question in a group does not depend on answering any other question.

No knowledge of formal logic is required for solving these problems. Type II questions are intended to be answered using knowledge, skill, and reasoning ability which are expected of college students and graduates.

Each group of questions is based on a set of conditions that establish relationships among persons, places, things, or events. The relationships are common ones such as temporal order (X arrived before Y but after Z); spatial order (H always sits in front of G and behind F); group membership (If Professor Green serves on the committee, then Professor Brown must also serve); and family structure (Mary is Juan's mother and Belinda's sister). The conditions should be read carefully to determine the exact nature of the relationships involved. Some relationships are fixed (P and R always sit at the same table). Other relationships are variable (O must be assigned to either table 1 or table 3). Some relationships that are not stated in the conditions can be deduced from those that are stated. (If one condition about books on a shelf specifies that Book L is to the left of Book Y, and another specifies that Book P is to the left of Book L, then it can be deduced that Book P is to the left of Book Y.)

Suggested Approach

¶Some examinees prefer to answer first those questions in a group that seem to pose little difficulty and then to return to those that seem troublesome. It is best not to start one group before finishing another because much time can be lost in returning to a question group and reestablishing familiarity with its relationships. Do not assume that, because the conditions for a set look long or complicated, the questions based on those conditions will be especially difficult.

¶In reading the conditions, do not introduce unwarranted assumptions; for instance, in a set establishing relationship of height and weight among the members of a team, do not assume that a person who is taller than another person must weigh more than that person.

¶It is intended that the conditions be as clear as possible; do not interpret them as if they were designed to trick you. For example, if a question asks how many people could be eligible to serve on a committee, consider only those people named in the explanatory material unless directed otherwise. When in doubt, read the conditions in their most obvious sense. However, the language in the conditions is intended to be read for precise meaning. It is essential, for instance, to pay particular attention to words that describe or limit relationships, such as only, exactly, never, always, must be, cannot be, and the like. The result of the careful reading described above should be a clear picture of the structure of relationships involved, including what kinds of relationships are permitted, who or what the participants in the relationships are, and what is and is not known about the structure of the relationships. For instance, following a careful reading it can often be determined whether only a single configuration of relationships is permitted by the conditions or whether alternative configurations are permitted.

¶Many examinees find it useful to underline key points in the conditions or to draw a diagram representing the configuration, as the directions for the section suggests.

¶Even though some people find diagrams to be very helpful, other people seldom use them. And among those who do regularly use diagrams in solving these problems, there is by no means universal agreement on which kind of diagram is best for which problem or in which cases a diagram is most useful. Therefore, do not be concerned if a particular problem in the test seems to be best approached without the use of diagrams.

¶Each question should be considered separately from the other questions in its group; no information, except what is given in the original conditions, should be carried over from one question to another. In some cases a question will simply ask for conclusions to be drawn from the conditions as originally given. An individual question can, however, add information to the original conditions or temporarily suspend one of the original conditions for the purpose of that question only. For example, if Question #1 adds the information "if P is sitting at table 2,...," this information should NOT be carried over to any other question in the group.

Sample Questions

Each group of questions is based on a set of conditions. In answering some of the questions it may be useful to draw a rough diagram. Choose the best answer for each question and blacken the corresponding space on your answer sheet.

Questions 1-3

Six people are standing in a ticket line, one behind the other, all facing the ticket window.

Each person is wearing a colored hat.

There are two people wearing blue hats, two people wearing red hats, one person wearing a green hat, and one person wearing a yellow hat.

The number of people between the two people wearing red hats is the same as the number of people between the two people wearing blue hats. That number may be zero.

1. All of the following are possible color arrangements for the first three hats, in order from first to third, EXCEPT

(A) blue, yellow, red
(B) red, yellow, blue
(C) red, red, blue
(D) yellow, blue, green
(E) yellow, green, blue

The answer to this moderately difficult question can be determined as follows: examine each answer choice and decide whether it is possible to complete the arrangement of the hats in a way that would not violate any of the conditions.

In choice (A), the arrangement BYRBGR

is possible (with B standing for a person wearing a blue hat, Y for a person wearing yellow, and so forth.) This rules out (A) as a possible answer, since the question asks you to identify the arrangement that is NOT possible.

In choice (B), RYBRGB is possible, which rules out (B) as a credited answer.

In choice (C), RRBBGY is possible, which rules out (C) as a credited answer.

In choice (E), the arrangement YGBBRR is possible, which rules out (E) as the credited answer.

In choice (D) however, no acceptable order can be constructed. The first three hats are YBG, leaving one blue hat and two red hats to complete the line. An attempt to place the remaining blue hat in each of the remaining positions (fourth, fifth, and sixth) shows that no acceptable order exists:

YBGBRR
or
YBGRBR
or
YBGRRB

In each arrangement, there would be an unequal number of people between the people wearing the two blue hats and the people wearing the two red hats. Choice (D) is the credited answer, since it is the only arrangement that is NOT possible.

The explanation of the way to answer this question demonstrates several points that may be useful in answering other questions of this type. First, a simple diagraming system in which B stands for a person wearing a blue hat, Y stands for a person wearing a yellow hat, and so forth, makes it easy to diagram each order in a concise manner. Second, in this question, which asks you to determine which is not a possible arrangement, it is necessary to determine only one possible arrangement to rule out each of the offered choices, even though there might be other possible arrangements that would satisfy the conditions. Third, in a question that asks you to determine which choice is NOT acceptable, it is very important to keep in mind that what you are seeking as the answer is an UNACCEPTABLE combination.

2. If the fourth person is wearing a yellow hat, which of the following is a complete and accurate list of all of the people who could possibly be wearing the green hat?

(A) The third person
(B) The first and third persons
(C) The first and sixth persons
(D) The first, second, and third persons
(E) The first, third, and fifth persons

This moderately difficult question can be approached by considering each of the choices in turn.

In choice (A) it is relatively simple to determine that the third person could be wearing the green hat. The arrangement BBGYRR satisfies all the conditions. This makes choice (A) attractive, but the careful test-taker will notice that the question asks for a "complete" list of the people who could possibly be wearing the green hat; therefore it is necessary to continue.

In (B), it can be demonstrated that the first person could be wearing the green hat. For example the arrangements GBBYRR satisfies all the conditions. This demonstrates that choice (B) is a possible answer, and also rules out choice (A) since (A) is not a complete list of the people who could be wearing the green hat.

In choice (C), it is not necessary to make a diagram. Since it has already been shown that the third person could be wearing the green hat, (C) cannot be a "complete" list of all the persons who could be wearing it.

In choice (D) it is necessary to determine whether or not the second person could be wearing the green hat. One way to do this would be with the following diagram:

1	2	3	4	5	6
	G		Y		

This shows that no matter how the red hats and the blue hats are distributed, there will not be an equal number of people between the two wearing the blue hats and two wearing the red hats. This rules out choice (D) as the credited answer, since the question asks for "a complete and accurate" list of the people who could be wearing the green hat.

In choice (E) a similar diagram could be constructed:

1	2	3	4	5	6
			Y	G	

This also shows that no matter how the red hats and the blue hats are distributed, the conditions cannot be satisfied.

Choice (B) therefore, is the credited answer. Notice that in order to answer this question, it is not necessary to determine that the sixth person cannot be wearing the green hat, nor is it necessary to determine all of the different acceptable arrangements. It is, however, very important to pay close attention to the wording of the question, to be sure that the answer chosen is the "complete and accurate" list of all of the people who could be wearing the green hat.

3. If both people wearing red hats are farther back in line than both people wearing blue hats, and if there is at least one person between the person wearing the yellow hat and the person wearing the green hat, which of the following must be true?

    I. The third person is wearing a blue hat.

    II. If the first person is wearing a blue hat, the fourth person is wearing a red hat.

    III. The third and fourth persons are wearing hats of different colors.

    (A) I only
    (B) II only
    (C) III only
    (D) I and III only
    (E) II and III only

One approach to this difficult question would be to construct a diagram showing all of the possibilities allowed under the restrictions. From line (i) it can be determined that I need not be true. Lines (i) and (iv) show that II must be true. The diagram as a whole shows that III must be true. Therefore, (E) is the answer to this question.

Question [illegible]	1	2	3	4	5	6
(i)	B	B	G or Y	R	R	G or Y
(ii)	G or Y	B	B	G or Y	R	R
(iii)	G or Y	B	B	R	R	G or Y
(iv)	B	G or Y	B	R	G or Y	R

Question Type III

This type of question is designed to evaluate both reading and reasoning skills. With respect to reading, the questions require an accurate and detailed understanding of the situation established by 1) a descriptive set of facts that leads to a dispute or case and, 2) two rules that can be used to govern the outcome of the dispute or case.

With respect to reasoning, the questions require an ability to identify relevant information and recognize how to apply that information in a sequence of steps. Specifically, this will require the ability to: recognize when or under what conditions a general rule should be applied to a specific set of facts; recognize when the application of two rules to a dispute or case will lead to different outcomes; determine whether new information is relevant to the application of a general rule to a specific set of facts; determine how the issue raised by a question bears on the application of a rule to a given set of facts.

Although the facts and the rules may be real or imaginary, they do not presuppose knowledge of law. The rules may be written using language similar to that used by advocates and jurists, but no actual knowledge of law is needed to understand the rules. If it is necessary to know the meanings of any technical terms that are present, their meanings will be explained. It must be kept in mind that though these problems may seem like those dealt with by the law, they are tests of reading and reasoning abilities only.

All of the questions of this type are presented in sets. Each question in a set must be evaluated independently of all other questions in the set. Each question and the issue that it raises must be examined only in terms of the facts and the rules given for that set.

The basic elements of these problems are:

The Facts:	a factual description of the characteristics, the specific actions, and the relationships between people or institutions;
The Dispute or Case:	a conflict that arises between the people or institutions decribed in the facts;
The Rules:	the two rules that, when applied to the facts either independently or together, govern the outcome of the dispute or case;
The Questions:	the issues to be classified in terms of their importance in applying the rules to resolve the dispute or case.

The exercise in these problems is one of classification. The issues raised must be classified in terms of their relationship to the application of the rules to the dispute or case described in the facts. Issues can be classified in four ways.

(A) A major issue whose resolution requires a choice between the rules.

Issues of this type require the application of the rules given to govern the outcomes of the dispute or case. If the application of each of the rules leads to a different outcome, and there is no basis for making a decision about which rule to apply, a choice between the rules must be made.

(B) A major issue whose resolution requires additional facts or rules but does not require a choice between the rules.

Issues of this type require further information that will establish important facts, clarify important terms in one or both of the rules, and/or provide additional rules relevant to deciding the outcome. However, these issues do not require a choice between the rules; they establish whether a major condition of a rule is met.

(C) A major issue that is readily answered by logical reasoning or by a reasoned application of the rules.

Issues of this type can be determined by understanding the facts or by the application of the rules to the facts. This determination requires only the careful use of common sense.

(D) An issue that is not logically relevant to the dispute or case at the end of the facts or that raises remote and unlikely possibilities.

Issues of this type, though they may be significant in other contexts, do not bear on the outcome of the dispute or case as governed by the given rules.

Some questions only raise issues; some questions supply new information as about facts and/or rules as well. Each question must, however, be considered independently, in terms of any new information the question may contain and the given set of facts and rules.

Do not assume that there will be a single right outcome for the dispute or case. The rules can conflict with what seems right or just to you. Preconceived notions of how the law works or ought to work should not be considered in classifying the issues. The issues must be classified on the basis of what is stated in the facts and the rules, and what may be logically inferred from the facts and the rules, and common sense.

Suggested Approach

¶Read the facts carefully. Determine accurately and completely the kinds of relationships that are established between or among the people or institutions described. Determine what the dispute or case is about, who the parties involved in the dispute or case are, and how the dispute or case defines the conflict.

¶Read the rules carefully. Determine what requirements and responsibilities are established. Determine specifically and in detail what the differences are between the two rules. Pay particular attention to such words as if, and, or, all, cannot, and only when. These words, and words like them, are very important in establishing how the rules are related to each other and how the rules will govern the outcome of the dispute or case.

¶Correlate the terms, requirements, and responsibilities established by the rules with the persons, institutions, and actions given in the facts.

¶Read each question carefully. The task is to identify the issue.

¶Ask the following questions: Does the issue concern information that is relevant and important to the dispute or case? If it does not, classify the issue (D).

¶If the issue is relevant and important to the dispute or case, ask the following question: Does the issue make it necessary to choose between the two rules, both of which, as far as can be determined, apply to the case? In these problems, you must not automatically suppose that one of the two rules takes precedence over the other. You must accept that the rules are of equal importance. If you must choose between the given rules to resolve the issue, then classify the issue (A).

¶Does the issue concern missing information, such as facts or rules, that is important in determining how to apply the given rules but that does not require you to decide which of the rules takes precedence? Classify such issues (B).

¶Does the issue concern matters that can adequately be determined by understanding the facts, by logical inferences from the facts, or by a logical and justified application of the rules to the facts? If so, classify the issue (C).

Sample Questions

Each of the sets in this section contains a statement of facts and two rules. These rules may be either real or imaginary, but for purposes of this test you are to assume them to be valid. The rules are followed by questions raising issues that may pertain to the application of one or both of the rules to the dispute or case presented at the end of the facts. Classify the issue raised by each question according to the following choices and blacken the corresponding space on the answer sheet.

(A) A major issue whose resolution requires a choice between the rules

(B) A major issue whose resolution requires additional facts or rules but does not require a choice between the rules

(C) A major issue that is readily answered by logical reasoning or by a reasoned application of the rules

(D) An issue that is not logically relevant to the dispute or case at the end of the facts or that raises remote and unlikely possibilities

Facts: Andy and John had been partners for twenty years when Andy's wife, Beth, divorced him and married John. Andy decided to get even with John by arranging the theft of a gold necklace that John had given to Beth. Andy was afraid that if he tried to take the necklace himself he would be recognized, so he told Oscar, his younger brother, that if Oscar did not steal the necklace and keep it for himself, he, Andy, would kill him. Andy knew that he could not kill his brother, but Oscar, who had been afraid of Andy since childhood, was frightened and so he decided to take the necklace. The night that Oscar took the necklace from Beth's dresser, Beth wakened and recognized Oscar. Oscar was charged with the crime of grand larceny.

Rules: I. One who takes personal property with a value of over one hundred dollars from the possession of another with the intent to steal the property is guilty of the crime of grand larceny.

II. A person who commits what would otherwise be a crime concerning property because he believes that his life will be endangered if he does not do so has not committed a crime.

1. How did Oscar get into the home of John and Beth?

2. If the necklace was worth five hundred dollars and if Oscar intended to steal the necklace because he believed that he would be killed if he did not do as Andy said, has Oscar committed grand larceny?

3. If the necklace was worth eighty-five dollars and if Oscar believed that Andy would kill him if he, Oscar, did not steal it, has Oscar committed grand larceny?

4. What was the value of the necklace taken from Beth's dresser?

1. How did Oscar get into the home of John and Beth?

This very easy question focuses on the distinction between a major issue and an issue that is not logically relevant to the case. Oscar is being charged with grand larceny, and the two rules establish conditions that determine whether or not grand larceny has taken place. The conditions have nothing to do with the way in which Oscar entered the house. Thus, the credited answer is (D).

2. If the necklace was worth five hundred dollars and if Oscar intended to steal the necklace because he believed that he would be killed if he did not do as Andy said, has Oscar committed grand larceny?

In order to classify the issue raised by this question, you must first apply the new information in the question to the situation created by the facts and the two rules. In order for Rule I to apply, two conditions have to be met: first, the necklace has to be worth more than one hundred dollars, and second, the person who takes the property has to take the property with the intention of stealing it. If the necklace was worth five hundred dollars and if Oscar intended to steal the necklace, then Rule I does apply. If Oscar believed that he would be killed if he did not do as Andy said, then Rule II also applies to the case. Note that both rules apply but that there is no logical basis for choosing one rule over the other. Moreover, the outcome would be different under the two rules: the issue "has Oscar committed grand larceny?" is a major issue that cannot be resolved without choosing between Rule I and II. The credited answer to this question of moderate difficulty is (A).

3. If the necklace was worth eighty-five dollars and if Oscar believed that Andy would kill him if he, Oscar, did not steal it, has Oscar committed grand larceny?

The issue raised by this question is one that can be answered by logical reasoning. If the necklace is worth less than one hundred dollars, then Rule I does not apply, regardless of whether Oscar intended to steal the necklace. If Oscar believed that he would be killed if he did not take the necklace, then Rule II does apply. The conditions that would make Oscar guilty of grand larceny do not apply to the case, while the conditions that would establish his innocence do apply; therefore, Oscar has not committed grand larceny as it is defined by the rules. The issue raised, then, should be classified as (C). This is a fairly difficult question.

4. What was the value of the necklace taken from Beth's dresser?

Before classifying the issue raised by this question, you should review the facts. The facts state that

Beth's necklace is gold, which suggests that the necklace might be of great value. However, since it is not known how much gold is in the necklace or what condition the necklace is in, it is impossible to infer the necklace's value from the facts. The value of the necklace, however, is a major issue, since Rule I applies only if the value of the necklace exceeds one hundred dollars. The credited answer to this question of average difficulty is (B). Note that the classification of the issue raised by this question should be made independently of the new information supposed in questions 2 and 3.

Question Type IV

The purpose of this type of question is to evaluate aptitude for understanding, analyzing, using, and critizing a variety of arguments. The questions are based on short arguments that are drawn from philosophical treatises, literary works and criticism, and other materials from the humanities, from the social sciences, from scientific articles, and from sources such as political speeches, advertisements, and informal discussions or conversations.

The questions sample a variety of abilities that can be considered subtypes of the ability to reason logically and critically. Questions measuring this ability require that you be able to (1) recognize the point of an argument, (2) perceive presuppositions essential to or supporting an argument or chain of reasoning, (3) draw conclusions from given evidence or premises, (4) infer missing material (such as implied arguments or antecedent and follow-up statements), (5) apply principles that govern one argument to another argument, (6) identify methods of argument and persuasion, (7) evaluate arguments, (8) differentiate between statements of fact and opinion, (9) analyse evidence and (10) assess claims critically.

It should be emphasized that the questions do not presuppose knowledge of the terminology of formal logic. You would not be asked, for example, to evaluate an argument by describing it as argumentum ad hominem but would be expected to recognize the unreasonableness of an attack on a person when an attack on an idea is more appropriate.

Suggested Approach

¶Read each question carefully. This type of question does not require any formal background in logic. All that each question requires is a close and careful evaluation of the reasonableness of the material given. This evaluation can be based completely on common sense.

¶Make sure that you understand the meaning of each part of the question. What relationships are established? What steps in the arguments are given? Are any seemingly necessary relationships omitted? Are any steps in the argument omitted?

¶Read each choice carefully. Remember that you are basically evaluating the reasonableness of a given argument or inference or being asked to reach reasonable conclusions from a given argument or inference.

¶Make sure that you understand the meaning of each choice and the ways in which it may or may not relate logically to the question asked.

¶This type of question does not involve any tricks or hidden meanings. Simply apply common sense to each question.

Sample Questions

The questions in this section require you to evaluate the reasoning contained in brief statements or passages. In some questions, each of the choices is a conceivable solution to the particular problem posed. However, you are to select the one that is best, that is, the one that does not require you to make what are by common-sense standards implausible, superfluous, or incompatible assumptions. After you have chosen the best answer, blacken the corresponding space on the answer sheet.

1. Since all rabbits that I have seen have short tails, all rabbits probably have short tails.

   Which of the following most closely parallels the kind of reasoning used in the sentence above?

   (A) Since all chemical reactions that I have seen have been undramatic, probably only minor changes took place in the substances involved.
   (B) Since all the human social systems that I have heard of have sexual taboos, all of these sexual taboos have probably had survival value for the human race.
   (C) Since all of the plays of Jovita Maldonado that I have seen feature a spruned lover, probably all of her plays feature this character type.
   (D) Since all eating utensils that I have seen are made of metal, metal is probably the most desirable material for eating utensils.
   (E) Since sight is the most important of human's five major sense, its failure probably seriously affects an individual's aptitude for all formal education.

The given statement reflects inductive reasoning, that is, a generalization about an entire class based on specific observations. The conclusion reached could be criticized because of the limited number of observations, but the question asks only for a recognition of a parallel example of the same kind of reasoning. The question is quite easy.

All the answer choices are similar in some ways, but only one is a statement about specific observations followed by a generalization based on those observations. In (A), (B), and (D), the second part of the statement is not a generalization based on the observations mentioned in the first part but is an explanation or suggested reason for what was observed. In (E) an assumption is followed by a conclusion. Only (C) refers to specific observations (about some of Maldonado's plays) and proceeds to generalize (about all of Maldonado's plays) on the basis of those observations, however limited they may be. Therefore (C) is the credited answer.

2. A good hotel can give you a beautiful room for $30 a day, with three meals, and make a profit and pay taxes. And yet a tax-exempt hospital operates in the red for $65 a day. I say it must be bad administration. The author's argument would be considerably weakened if attention were drawn to the fact that:

   (A) hotel managers receive better training than do hospital administrators
   (B) the quality of food served by hotels exceeds that of food served in hospitals
   (C) hospitals are run by dishonest administrators
   (D) hospitals provide other services besides room and board
   (E) hospitals deficits are a recent phenomenon

This very easy question focuses on the reasonableness of drawing a conclusion from the evidence presented. The author's contention is based on some evidence--the discrepancy between a hotel's and a hospital's operation expenses. The question asks you to identify additional evidence that would weaken the argument that bad administration is responsible for the discrepancy in expenses. (A) and (B) cannot be that evidence because these choices, if true, would actually strengthen the author's argument. (C) is slightly altered version of the author's own statement of a reason for the discrepancy. (E) could weaken the argument, but only if more information were given. Only (D) provides evidence that casts doubt on the argument. If hospitals provide services other than those mentioned, then the costs of those services rather than bad administration are likely to be the reason for the difference between hotel and hospital expenses. Therefore (D) is the credited answer.

Questions 1-4 refer to missing portions of the following passage. For each question, choose the completion that is best according to the context of the passage.

If a book disgusted everyone, no one would read it. However, one can be sure of selling many copies of a book that is publicly proclaimed obscene, for the officially held standards of propriety do not prevail throughout the community. At this point I may be expected to denounce the hypocrisy of the age. I shall not do so. The concept of hypocrisy applies to morals: a person should be good and not merely seem so, and a bad person is little mended by pretense of goodness. But propriety is altogether a matter of how actions appear, so that 1. If a person seems to give no offense, he gives no offense. Why, then, should a society not have public standards of propriety different from those applied by each citizen to his own private conduct? It would be no more absurd to advertise filthy movies by decorous posters than it is to advertise decorous movies by filthy posters; and if a society in which everyone avidly reads pornography were to forbid its public sale, that would mean only that it combined a taste for such reading with a taste for 4.

1. (A) bad men rarely succeeded in appearing good
   (B) the concept of hypocrisy does not apply
   (C) actions appear different to observers with different standards
   (D) the issue is basically a moral one
   (E) the concepts of impropriety and immorality are indistinguishable

Both this question and the next require that you follow the author's reasoning well enough to fill in missing material. The first question focuses on the distinction that the author makes between morality and propriety. The author suggests that propriety, unlike morality, is entirely a matter of appearance and has nothing to do with what is really good or bad. Because the author contrasts morality and prioriety, (D) and (E) can be eliminated. (D) assumes that propriety is a moral issue. (E) suggests that impropriety and immorality (and by inference propriety and morality) are indistinguishable. Since "bad" and "good" in (A) refers to morality, (A) does not follow from the author's idea on propriety. (C) is an appealing answer, since it focuses on the way actions appear in public. But (C) does not follow from the words "so that," leading from the first clause ("propriety is altogether a matter of how actions appear") to the second clause. Only (B) is an acceptable answer. It follows from the author's distinction between propriety and morality because it states that hypocrisy (associated with the concept of morality) does not apply to the question of propriety. This question is difficult.

4. (A) obscenity
   (B) hypocrisy
   (C) oppression
   (D) good art
   (E) decorum

The answer to this very difficult question must meet two requirements. It must describe something that is consistent, in the context of the author's discussion, with the reading of pornography. It must also explain the paradox, described in the preceding lines, of a society in which pornography is read although public sales of pornography is forbidden. The sense of the author's argument is that it is possible not to call such a society hypocritical (lines 4-8). Therefore (B) a is a poor choice. (A), "obscenity," is a term similar to "pornography" and does not suggest any contrast that could explain the paradox. "Oppression" and "good art," though relevant in general to the topic of pornography, are not relevant to the author's discussion here. Thus (C) and (D) can be eliminated. (E), "decorum," a synonym of "propriety," does fit into the context. It explains the contrast between privately reading pornography and publicly banning it, according to the author's view that propriety has to do with public appearance only and not with private actions. Thus (E) is the credited answer.

Mr. WEISS. Thank you, Mr. Zimmer.

Mr. Erdahl?

Mr. ERDAHL. Thank you very much, Mr. Chairman.

Thanks to both of you for what I think has been an especially reasonable and studied approach, almost conciliatory, we might say.

We heard, Mr. Zimmer, you are telling us go slow and see what happens in the New York experience.

Dr. Hanford, would that be your similar advice?

Mr. HANFORD. Yes, sir, it would.

Mr. ERDAHL. Also, as we look and read and listen to the testimony of both of you, it seems apparent that you are underscoring what others have told us today and in past times, that some type of selection process is bound to be used, the exact alternative or complete opposite alternative would be to put everybody's name in a hat and draw out a number.

But I gather from what both of you have said that these tests are a supplement to other things that go into the selection of a student, not so much to help the institution to select students that might have an ability to succeed and achieve, but it also seems to me that the tests are a very valid tool that identify applicants, and that students can use to determine, if they should go to St. Olaf College or to the University of Minnesota.

Can you comment on that? We have often approached this from how they are used by the institutions and by the testing agencies, but could each of you elaborate a bit as you see the use of tests for students?

Mr. HANFORD. I mentioned in my comments that when I joined the board students were not allowed to know their scores. That decision was reversed by the members of the board in 1957, and the minute we began making the test scores available, we committed ourselves to a process of guidance in connection with college admissions.

We developed the preliminary scholastic aptitude test as a guidance tool for young people to use. Sure, practice before you take the SAT test for real.

I think you will find, jumping ahead 25 years, in our most recent literature, that the college admissions process is no longer as competitive a process that it was some years ago, that it is more a matching process and that new testing programs, and new services that we are trying to develop, are as much if not more to help young people make decisions about their higher education opportunities than to help institutions make choices among candidates. There aren't very many of them that have that opportunity.

Mr. **Erdahl**. Thank you.

Mr. **Zimmer**. There is a slightly different coloration in our case because of the fact that law school admission remains intensely competitive, only slightly more than 60 percent of those persons applying to the ABA credited law schools last year were admitted.

I think that, so that our test has rather more force in the fair allocation of those places than it has in the guidance of many people. I think it helps people, particularly at the margins, developing disadvantaged people in particular.

That is the greatest use as a diagnostic device in determining readiness for legal study, but for very many of our people it is operating far above the threshold of minimum competence, so it is helping the individuals select a law school at which they would do best and it is helping all of the law schools to allocate their places in the fairest fashion.

Mr. **Erdahl**. Thank you very much and again thanks for some very well thought out testimony.

Thank you, Mr. Chairman.

Mr. **Weiss**. Mr. Zimmer, let me see if we can follow up a little bit on the last question Mr. Erdahl asked, that in this context of continued competitiveness for positions within the law schools, the Law School Admissions Council has decided that the validity of its tests would not be compromised by full total disclosure; is that right?

Mr. **Zimmer**. I don't think we can know that in advance. I think that this is a situation not free of ambiguity. It is a question of trading off values in a rather complex way. We as lawyers and persons who are primarily consumers of tests have had a great deal invested in our own value structure in the concept of openness and it was something we brought a lot to.

We are doing everything we can and our test developers are doing what they can to assure that there won't be a compromise in validity, but I don't know that I would tell you that if I thought there was a compromise of one or two numbers in the hundreds or thousands places, that I wouldn't be willing to make that tradeoff.

Mr. **Weiss**. The decision that you have made, that is, the council has made, is to disclose automatically now the test results and the questions to all those who take the law school admission test; is that right?

Mr. **Zimmer**. Yes, sir.

Mr. **Weiss**. So that when you are saying wait and see, you yourselves have in fact gone beyond the waiting and the seeing, you have come to a judgment and are following through on action on that judgment and making full disclosures?

Mr. **Zimmer**. In our program with proven high demands for this service, much of the decision is a practical business decision, that to respond to and answer orders from 60 percent of the people is, for example, more costly than simply to send it out automatically, so at a certain level there is a simple business judgment, a test of the marketplace in our area which has shown that law school applicants in a competitive environment want this service has led us to make that decision.

Mr. **Weiss**. Dr. Hanford, you have indicated that you have had a feedback of only 3 percent of those who take the SAT or took it this past year asking for their test results.

Mr. **Hanford**. That was the average in 1981, yes.

Mr. **Weiss**. Right; now, I understand, and you two gentlemen correct me if I am wrong, that there is a difference in the form, the method of request for disclosure, between the LSAT's and the SAT's.

Would you each explain how a person has to request their test results and the questions?

Mr. **Zimmer**. There is no request necessary. It is an automatic part of our service.

Last year it was a part of our registration form, a person could designate the request for this service by checking a box on the application form and paying the fee.

Mr. **Hanford**. In our program we have a National Bulletin that goes to all students to register for the SAT, and it has at the bottom special information for students who plan to register in New York.

We have to have a special section because of the particular requirements of the La Valle Act, in any event. There is a form which is right here which is no more complicated to fill out than the registration form itself, right next to the envelope in which you send it with a description of the procedures for requesting your SAT questions.

The law schools last year had a checkoff; this year it is automatic. Ours is a form that the student has to write his or her name and address and submit it.

Mr. **Weiss**. A separate form?

Mr. **Hanford**. A separate form, right.

Mr. **Weiss**. It doesn't go in the same envelope as the registration?

Mr. **Hanford**. Right here; the same.

Mr. **Weiss**. Is it the same envelope in which the person files the registration?

Mr. **Hanford**. It can be; it doesn't have to be.

Mr. **Weiss**. Is there a box on the envelope which says check this off if you want your test results?

Mr. **Hanford**. No.

Mr. **Weiss**. Has any thought been given to making it really simple for an applicant to check off a box?

Mr. **Hanford**. We thought very seriously about it. We felt it was important to test the demand.

Second, we still question the educational value of disclosure. One of the telling comments made to me by Senator La Valle in talking about his son who wasn't much interested in the law until he took the tests, and when the scores came he was kind of curious as to where he had gone wrong. For that reason, if someone wants to know where they made a mistake, this kind of service is available, but we didn't feel that it was educationally sound.

We still don't believe disclosure is educationally useful. For that reason we felt it was perfectly appropriate.

Mr. **Weiss**. Are you trying to discourage people from asking?

Mr. **Hanford**. No, sir; I do not think you will find this format and procedure discouraging. I think any young student in New York today is aware from the bulletin, from the press, from a guidance counselor, that this service is available. There is a real difference between our situation and the law schools.

By the time the young person sits down to take the SAT, that young person has probably had a chance to take what we call the PSAT, which is a practice test for taking the SAT. Then they receive "Taking the SAT" and then they take the SAT itself. So they have three cracks at a test like the SAT and I think that 3 percent is probably a reasonable measure of the demand.

Mr. **Weiss**. But you don't really want to find out for sure by making it very simple for them to ask.

Mr. **Hanford**. We don't think at this point that would be educationally sound.

Mr. **Weiss**. Well, I must tell you, I am impressed and at the same time sort of disappointed by ETS and the College Board. I think that it is true that the College Board and the Educational Testing Service people have come a long, long way in a very few short years, and that is because for being proud on your part that you are responsive to what American society seems to be asking of you.

At the same time, I think that your attitude in this particular instance sort of typifies how you came to the point of which you are so proud now, you came more or less being dragged by your heels, when, in fact, you had no choice anymore, you said OK, you will do it and that seems to be what you are doing at this point and in a couple of years you will probably come around to at least making it simple for people to ask for their forms and take great pride in the fact that you are doing it, even though you are coming to it with the most extreme reluctance, and I don't understand why, because it undercuts some of the very major accomplishments that you have achieved in the last couple of years.

You may comment on that but that is the way it seems to me sitting on this side of the table.

Mr. **Hanford**. It's a very honest difference of opinion, Mr. Chairman. We feel we have made an honest effort here; I still don't believe that anybody can say we have been trying to hide anything. I think 3 percent is a perfectly reasonable response, a good measure of interest on the part of the students, for reasons I have suggested, and all the more reason not to try to con some kids into buying a service they don't really need or want by providing a checkoff.

Mr. **Weiss**. Do you have any information on any of the other postsecondary institutions admissions tests, higher education institutions admissions test results on any of the graduate schools?

Mr. **Hanford**. The College Board is responsible and concerned only with the tests for undergraduates. While I know something about the others, I should not comment. It would be inappropriate.

Mr. **Weiss**. OK. Are you familiar at all with what the American Council on Testing has been doing and how at all does that fit in with any of your programs as far as their disclosure policies are concerned?

Mr. **Hanford**. The American College Testing program?

Mr. **Weiss**. I am sorry.

Mr. **Hanford**. I am really unfamiliar.

Mr. **Weiss**. Information that we have is that in fact they have up to this point refused to disclose at all, and in essence it is part of the problem that we face in a situation where some of the testing companies have gone forward at least a significant portion of the way and then they say, well, wait and see when in fact there are still large portions of our set out there which have tests administered by other people where there is no disclosure.

Mr. **Hanford**. You point to another interesting difference between the law school circumstance and the College Board circumstance which is that people occasionally tend to think of the College Board or Educational Testing Service as a monoply, and we do have competition.

Mr. **Weiss**. Limited though it may be.

Mr. **Zimmer**. We are for openness but we don't tolerate competition.

Mr. **Weiss**. I gather, Mr. Hanford, that there has been some change now in the attitude of the College Board as to the effectiveness, if not desirability, of coaching, of preparation schools.

Could you comment on that?

Mr. **Hanford**. I would like to associate myself with the comments that Dr. Johnson made earlier. I tend to put it in my own context. It all depends on how you define coaching on the one hand and instruction on the other, and the problem gets to be a semantic one in the middle.

When does coaching, cramming, drill on item types stop being cramming and coaching and when does it become good instruction?

Our statements over the years, I think, have been consistent if one takes the very narrow definition as we always did of coaching, as the drill practice and so forth on test items.

You called attention in your comments to the FTC study. I would invite the committee to read very carefully the caveats and demurrers that accompany both the initial report and the subsequent one that was just recently released, because they do suggest that there are problems having to do with the nature of the technical quality of the research, and with what is known as the self-selection factor.

The young people who go to commercial coaching schools tend to be students who are highly motivated perhaps and so please read carefully the demurrers that are at the beginning that cover the release of the latest FTC study.

We still believe that what we do for young people in getting ready for the PSAT and taking the SAT is all that young people really need to get themselves familiar with that type of instrument, that commercial coaching that emphasizes simply drill and so forth really does not work.

But when you get over toward the other end of the spectrum to studies that go on for a year's time you may have something else. After all, seniors do better than juniors on the SAT, and we expect them to because we expect the SAT scores to be influenced by education. That is what it is supposed to measure.

Mr. **Weiss**. Of course, as you know, part of the criticism that has been directed at testing agencies was that by creating the impression, perhaps unintended, that you could not really study for the SAT's that the people felt in fact it did not test what they were learning.

It tested some vague undefinable aptitude, if you will.

Mr. **Hanford**. One of the things that is happening in the psychometric world, and I should say to you both that I don't have that doctor in front and my training was in business. I am not a psychologist or psychometrician, but I know a little now about the field and I think we have tended to emphasize, we have not emphasized but, as you suggested the context in which we thought about the test has misled people to equate it with the old-fashioned IQ, which has never been our intention. It is in the record that this is not what we are talking about. Schooling has an effect, but I think in emphasizing that cramming, drill, and so forth, was not really good education and saying that it did not work, we may very inadvertently have left the impression that the SAT was not susceptible to education. But that simply could not be the fact because, again as I said, seniors always have done better than juniors because it is a test that responds to education.

Mr. **Weiss**. The other part of the universe we are discussing that the legislation address as itself to is not simply the educational benefits to the students who take the test, but of improving the validity and reliability of the tests themselves and the questions by making the test results, and the in-house studies, available to outside experts other than in-house experts and I think that was the area Dr. Mehrens was commenting on when he suggested he was not in favor of releasing it to the students or to the whole world, but he seemed to be in favor of having the field of expertise broadened.

I am wondering whether in fact the College Board sees merit in this disclosure and mandated disclosure so that in fact you would not just have an in-house group which all the intended objectivity in the world may not be as capable of picking up flaws or faults in its work product than if it were subjected to regular critiques by outside experts who are interested in the whole testing world and the consequences of testing.

Mr. **Hanford**. An inference that I draw from your question is that we have tended to be very secretive about the tests and avoided criticism of the SAT. In fact there has been a policy of making copies of the tests available to researchers so long as they agreed to maintain the security of the tests and not publish items, so to speak.

It has not been a matter of not making it available to researchers. In fact some years ago, the education editor of the New York Times did a piece on test items. We made them fully available to the people he wanted to have review them. So I have to say we have not been secretive in that sense and as I said earlier, there are 1,500 test items them out there which we have published since January 1980 that people have had a chance to look at.

Mr. **Zimmer**. If I may, I would like to add my suggestion that you survey the New York education commissioner to see how many requests have come from researchers to review the materials that have been filed there.

I think you would find that it is surprisingly few. There is a need here to focus on the aptness of the remedy for the perceived problem.

There are just not people breaking down the door of the New York commissioners to get stuff that is filed there.

Mr. **Weiss**. You notice we have had testimony in the course of the 2 years during which we have had the legislation under consideration, coming from people with knowledge in the field, as to the difficulties they have had in getting not the test results necessarily, but the in-house studies and surveys and critiques.

We had testimony which indicated people had to go to court to try to get some information. We had testimony from a couple of Harvard psychologists and sociologists in a panel we had a year ago in which we had indications that the ETS had in fact omitted certain studies from some of its conclusional reports.

We have Dr. Burrow's statement as to the difficulty of really getting trustworthy information of adequate supply. So there seems to be sufficient basis for our concluding that the people in the field feel less than happy about the openness of ways an academic field really and should, in fact, be subjected to the rigorous testing and challenging of other experts within that field.

I don't know what there is to be lost by becoming more open in that area.

Mr. **Hanford**. Let me say that you may recall our chairman, Fred Hargadon, said in his testimony before you 2 years ago that we would like every person to be able to take home the test after their administrations.

What you have to balance off, and what we are trying to do is balance off questions of quality, cost, fairness, and convenience for students.

As I said, if we have to disclose more than five SAT's, given the current state of the art, and the current availability of people, we will have to restrict the number of administrations down to something on that order.

It is a matter of training people to write items. It is a matter of pretesting items; it is a matter of putting the forms together. These are processes that take time. We can't do this overnight.

I just don't think that is what you want to do, to restrict the opportunity for young people to help themselves get into college.

Mr. **Weiss**. Of course I don't and nobody here does, but two things: First, Dr. Irby, who is in the audience, was good enough to invite a number of us down to ETS headquarters outside of Princeton—I guess about 1 year or 1½ years ago—and it was a very educational and informational experience for all of us.

One of the things that I learned was that some 80 percent of the questions that anyone in the administration asked were, in fact, not new material.

It is not as if we are asking people to start from square 1. They were able to use 95-percent old material.

The second thing is that apart from the benefit and the role of the individual students, what we are addressing ourselves to in part of the legislation is the sharing, the openness, if you will, of information that goes into the test makeup, the test validation, the critiques of experiences with the test—all the things which social scientists in the field would be interested in, not part of curiosity's sake, but so they can make the field of testing a more valid and more reliable and more accurate one.

Again, I don't understand why ETS college boards have been so reluctant to provide that kind of access.

Mr. **Hanford**. Mr. Chairman, I insist that we have been making studies available about the SAT. In my office, we have what we call a green Bible which describes the mechanism by which the SAT and achievement tests were put together.

It was published in 1970. We have been making information available regularly to young people, to colleges, about how tests should be used, that they should not be used alone, that they should be used with other criteria in the administrations process.

Mr. **Weiss**. Do you object to it being used as a cutoff?

Mr. **Hanford**. Not exclusively; no, sir. I think there are times in our society when it is necessary to use gross means to reduce a large pool to some size. But in most cases, I would certainly not tolerate it.

If I may, I would like to take a brief minute because the purposes as set forth in section 3 about test information—and it relates to Mr. Simon's questions—that I just want to cite for the record that we will respond in detail to his concerns and questions about the 10 provisions of section 3.

I simply say at this point that I note in the third article the distinction made between tests used in baccalaureate and postbaccalaureate admissions. As I read this, I am not sure whether you are talking about success in careers as related to the SAT.

It seems to me—and I would echo indeed Dr. Zimmer's point—that I know of very few studies in which success in careers has been measurable.

Mr. **Weiss**. The reason that is in there is because a number of the testing companies and tests, if you will, in their promotional materials suggest or state to the would-be test-taker as to the correlation, the positive and affirmative correlation, between the results of those tests and how they will do either in school or how they will do in their careers.

What we have said in here is that, "OK, friend, if you are going to make that kind of statement, prove it. Show it."

Mr. **Hanford**. I wanted to make sure it did not relate to the SAT because it would not seem to me to be applicable.

In 4(A), I just have to say that the expression of predictions as expressed as a percentage is not useful. We have encouraged institutions to try to have a validity study and make sure that they put the SAT in its proper balance and perspective. I would say the same thing as to (C).

In No. 5, it says, "the test agency will use the score"—I must point out that we don't use the score. Institutions use the score. But in general, my response to your questions in terms of section 3 is, that we feel we have been doing this and are doing it now.

Mr. **Weiss**. We look forward to your response and may I just indicate both to this panel and to the previous panel that we would need to have your written responses in to us within 10 days. The record will be kept open for that 10-day period and then will be closed so we can get it printed.

If there are no further questions, I want to express my appreciation on behalf of both of the subcommittees and we will go on and we will—we are again appreciative of your participation in the process.

The committee stands adjourned until tomorrow morning at 9:30.

[Whereupon, at 1:50 p.m., the subcommittee was adjourned, to reconvene at 9:30 a.m., Wednesday, July 22, 1981.]

[Material submitted for inclusion in the record follows:]

AUG 4 1981

MICHIGAN STATE UNIVERSITY

COLLEGE OF EDUCATION
DEPARTMENT OF COUNSELING AND EDUCATIONAL PSYCHOLOGY
EAST LANSING · MICHIGAN · 48824

July 27, 1981

Mr. Paul Simon
Committee on Education and Labor
2181 Rayburn House Office Building
Washington, DC 20515

Dear Mr. Simon:

When I testified on July 21 on H.R. 1662 you asked me to respond to Section 3 of that bill in writing. My response follows:

Section 3(a) (1) and (2)

These two subsections are unnecessary since every test agency already provides such materials.

Section 3(a) (3)

The first part of this section is unnecessary since such data are already provided. The subsection requesting the correlation between test scores and success in a career is very unwise. Tests, by and large, are not designed to predict success in a career. They are designed to predict success in an educational/training program. While it is true that there is an empirical relationship between admission scores and success in a program and there must be some logical relationship between success in a training program and success in a career (e.g. knowledge of anatonomy and success as a physician), it is extremely difficult to empirically show a relationship between the first and third variables. This is primarily due to the difficulties of defining and measuring career success (e.g. income does not define success as a physician).

Section 3(a) (4) (A)

To communicate prediction "expressed as a percentage" is essentially meaningless and certainly not comprehensible to most candidates. A good example of how misleading this can be is found in the Nairn/Nader report (see my critique of this report appended to my July 21 testimony).

Section 3(a) (4) (B)

This information, if gathered, would be misunderstood by many people. Weiss himself does not understand the meaning of such data (see his July 31, 1979 testimony) and he has been informed many times of its meaning in previous testimony.

MSU is an Affirmative Action/Equal Opportunity Institution

Section 3(a) (4) (C)

No one really knows this and, if so, one certainly would not interpret it as a percentage!

Section 3(a) (S)

The test agencies provide data to colleges where the data get used. A test agency simply does not know how every college will use the data.

Section 3(a) (6) to 3(a) (9)

These sections are not terribly bad nor at all necessary.

Section 3(a) (10)

Since I don't feel the section 5(a) (2) is advisable, I can't approve of this section either.

Section 3(b) and 3(c)

Not needed.

Obviously I am opposed to Section 3 just as I am to other sections of the bill. It simply does not appear to be written by a person with any expertise in measurement.

Thank you for requesting the above comments.

Sincerely,

William Mehrens
Professor
Measurement and Evaluation
462 Erickson Hall

WM:cd

The College Board
888 Seventh Avenue, New York, New York 10019
(212) 582-6210

Office of the President

July 31, 1981

Honorable Carl D. Perkins
Chairman
Subcommittee on Elementary, Secondary
and Vocational Education
Committee on Education and Labor
U. S. House of Representatives
Washington, D.C. 20515

Honorable Paul Simon
Chairman
Subcommittee on Postsecondary Education
Committee on Education and Labor
U. S. House of Representatives
Washington, D.C. 20515

Dear Sirs:

I want to thank you for the opportunity to testify before your subcommittees last week on HR 1662, the Educational Testing Act of 1981.

During the course of the testimony, Mr. Simon requested that witnesses later submit comments for the record on the provisions of Section 3 of HR 1662, which specifies the information test agencies would be required to provide to test takers and to educational institutions. Our commentary in response to this request is enclosed, indicating which types of information the College Board now provides and in what form and noting where the requirements of Section 3 would be impractical or inappropriate with respect to the SAT and other tests we sponsor.

The enclosed supplemental statement also clarifies our answer to a question raised during the hearing by Representative Weiss regarding the College Board's procedures for enabling students to see their test questions and answers.

I am by copy of this letter providing the same supplemental materials to Mr. Weiss for his information.

I hope this additional information is helpful. Please let me know if the College Board might be of further assistance.

Sincerely,

George Hanford

George H. Hanford
President

GHH/dce

Enclosures

cc: Rep. Ted Weiss

A nonprofit educational association serving students, schools, and colleges through programs designed to expand educational opportunity

Supplemental Statement for the Record*

The College Board

Comments on Section 3 of H.R. 1662

H.R. 1662 would require test agencies to provide certain information to test-takers and to educational institutions that receive test scores. Section 3(a) of the bill specifies the types of information which each test agency would be required to "provide to any test subject in clear and easily understandable language, along with the registration form for a test...:" Section 3(b) requires that the same information be provided to any institution "prior to or coincident with the first reporting of a test score or scores" for a test to the institution. Section 3(c) requires test agencies to notify students and institutions "if the test subject's score is delayed ten calendar days" beyond the date it is due to be mailed.

H.R. 1662 would apply to three testing programs sponsored by the College Board--the Scholastic Aptitude Test (SAT), the Achievement Tests, and the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT). The results of the SAT and Achievement Tests are used by colleges and universities as part of the admissions process for undergraduate programs. The PSAT/NMSQT is a guidance instrument that helps high-school juniors make plans for college and prepare for taking the SAT; it is also used as the qualifying test for students who wish to participate in the scholarship competition conducted by the National Merit Scholarship Corporation.

The College Board now publishes and widely distributes materials containing most of the information specified in Section 3 of H.R. 1662. Copies of most of these publications have already been submitted to the subcommittees; another set is enclosed with this statement.

---

*Hearings on H.R. 1662, the Educational Testing Act of 1981, before the Subcommittees on Elementary, Secondary and Vocational Education and Postsecondary Education, July 21-22, 1981

These regular program publications cover the purposes and content of the tests, score reporting procedures, information to aid in score interpretation, confidentiality of data, and description of special arrangements for handicapped students. Thus, information that would be required by Subsection 3(a) paragraphs 1,2,3,6,8, and 9 is incorporated in College Board materials provided routinely to test takers as well as test users (institutions). Information is also provided about the College Board's voluntary question-and-answer disclosure service, privacy of test scores, and the availability of procedures for appeal or review of a test score [Section 3(a)(10)]. In addition, the College Board already follows the procedures that include notification of delays in score reporting to test candidates and institutions as would be required by Section 3(c).

Not all of the publications containing this information, however, are provided to test candidates along with registration forms. (In fact, for the PSAT/NMSQT there is no student registration form, since the answer sheet completed at the time of testing gathers the necessary information from individuals. School officials merely order the number of tests they will need to satisfy the demand from interested students in their schools.) College Board practice has been to provide information at the time it is most likely to be useful to and of interest to test candidates: some before the test is given, some at the test administration, and some when scores are sent to students. H.R. 1662 would restrict the College Board's ability to continue to provide information in the most timely manner throughout the entire testing process.

Another general concern we have about Section 3 is simply practical. While there is no reason test candidates should not have access to any available material about College Board tests, a requirement that all of it be included in program publications ("...along with the registration form for a test...") creates a problem of size and volume. The main registration booklet for the SAT and Achievement Tests, the Student Bulletin, already attempts to convey a large amount of procedural, theoretical and background information. The current edition contains 15 pages of text as well as 15 pages of regional codes for reference. We have found that many test takers--the overwhelming majority of whom are high-school students--have some trouble reading through, let alone fully understanding, all of the material provided in this publication. A law of diminishing returns tends to operate: the more information the Board must incorporate in the text, especially information of a technical, legal, or excessively detailed nature, the less likely students will be to grasp it fully and the less inclined they will be to try. (A student intern who has spent the summer reviewing our publications has reached the same conclusion and has corroberated our sense that more is not always better.)

The College Board's goal is to explain tests to students as clearly, comprehensively, and concisely as possible. We continually review and revise our publications to try and make them more relevant and useful. Some of the requirements under paragraphs [illegible] and 7 of Section 3(a) of H.R. 1662 are examples of information that we do not currently provide, exactly as specified, to all test takers.

Although the College Board already does or could comply at least in spirit with most of the provisions of Section 3, the requirements of 3(a)(4) would be especially problematic and their enforcement would add little, if anything, to public understanding of the tests.

Section 3(a)(4)(A) requires publication of statements on "the extent, expressed as a percentage, to which the use of [the] test score improves the accuracy of predicting future grade point average, over and above all other information used." The College Board now describes in its publications the ways in which SAT scores are typically used by institutions, in combination with information about the student's high-school record, to help predict freshman grade point average. It would be misleading, however, to attempt to attribute a generalized percentage increase in predictive validity to SAT scores alone. Even if it were possible for the College Board to know to what extent SAT scores alone improved the forecasting of college grade point average at every postsecondary institution in the country, such data would vary widely from one institution to another. The College Board and ETS have published considerable information about the predictive validity of test scores, much of which is summarized in the enclosed booklet, Test Use and Validity.

Section 3(a)(4)(B) would require providing all test takers information about the statistical relationship of test scores and family income. Such data are widely available through other publications of the College Board and ETS (see the enclosed pamphlet on Test Scores and Family Income); however, we have serious reservations about the desirability of incorporating this material in routine program publications for distribution to all test takers. While scores are, on the average, higher for students from families with higher incomes, students from each income level obtain the full range of SAT scores. Nearly one-third of the students with family incomes below $6,000 rank in the top half of the total group in terms of SAT scores. One cannot forecast SAT scores from family income. Thus, the College Board avoids making statements which might discourage low-income students from aspiring to higher education or from taking the SAT for fear that their scores will be low.

Section 3(a)(4)(C) would require providing test takers with information on the "extent to which test-preparation courses improve test subjects' scores on average, expressed as a percentage." Over the years the College Board and ETS have published numerous studies on the effects of various types of special preparation and, more recently, the Board has distributed to students a statement about special preparation for the SAT. A copy of the most recent summary of research on this subject, The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research from the Fifties to the FTC, is enclosed. The text of the statement for students about special preparation is included on the back cover of Taking the SAT, which is provided to all students who plan to take this test.

Presumably the intent of Section 3(a)(4)(C) is to summarize all available data about special test preparation courses in a single number. Since such activities vary widely in their quality, intensity and duration, the average score change that might be attributable to test preparation courses will also differ. Furthermore, studies on the effects of special preparation cannot separate the actual effects of the course itself from the role played by motivation, interest, effort, and other individual attributes on any possible score change. Oversimplification of the results of the numerous studies on this topic would be misleading and a disservice to many test candidates who may conclude that any test preparation course will yield the stated average percentage increase.

Most of the above comments regarding Section 3 of H.R. 1662 relate specifically to the SAT; many similar concerns would apply to the Achievement Tests and the PSAT/NMSQT. Some of the required information as it relates to the latter two programs is not readily available or would not be particularly meaningful. For example, there is no recent reliable data on special preparation courses for the Achievement Tests.

In summary, the College Board routinely provides to students most of the information called for by Section 3 of H.R. 1662 at some point during the testing process, which extends from registration through receipt of scores. Other information, while not provided to test takers in exactly the form specified by H.R. 1662, is either covered in general in the regular student-oriented program publications, is in the public domain or is available on request.

Procedures for SAT Question-and-Answer Service

For the record, the College Board would also like to clarify the procedures under which students may request to see test questions and answers on the SAT, both in New York State and nationally. The New York State Standardized Testing Law requires that students taking the test in New York be given the option of disclosure. Beginning in the 1981-82 testing year, the College Board had voluntarily adopted a national SAT question-and-answer service at five administrations of the SAT, giving students the opportunity after taking the test to order copies of their answer sheets, the correct answers, and the questions that counted towards their SAT scores.

During this past year, New York students received a special edition of the Student Bulletin and a New York State Supplement which contained the order form for disclosure materials. (See page 7 of the enclosed 1980-81 New York State Supplement.) Because of the large number of students who take the SAT, it was necessary to have a procedure that would match a particular request for disclosure material with the correct records. The most reliable method to accomplish this was to have students provide their test registration numbers. Although students could send in their requests prior to receiving their scores, they did need to provide their registration number which they received on the admission ticket, the form sent to them upon receipt of their registration form. Incomplete requests, including those without this number, were generally returned to the test candidate. However, no major problems were encountered with this system and most students had no difficulty in completing the form as instructed.

For the national disclosure service of the SAT being introduced in 1981-82, the general system is similar. Large quantities (nearly four million copies) of the enclosed Special Announcement about the SAT Question-and-Answer Service are being shipped this summer to secondary schools for distribution to students along with the Student Bulletin. An order form is part of this announcement. The Board of Trustees of the College Board made the decision to offer this service after most of the revisions had been made in the regular program publications for 1981-82, therefore precluding incorporation of all this information in the Student Bulletin and other publications. It was possible, however, to call attention to the service by means of special notices on the covers of most of the publications. Staff are currently investigating other methods to offer the SAT Question-and-Answer Service to the more than 1.5 million candidates who are expected to take the SAT in 1982-83.

The Committee for Fair and Open Testing
5 Beckman Street, Room 1000, New York, N.Y. 10038 (212) 349-6460

CO-DIRECTORS:
John Weiss
Steve Solomon

TESTIMONY
of
STEVE SOLOMON
before the
U.S. HOUSE OF REPRESENTATIVES
SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION
and
SUBCOMMITTEE ON POSTSECONDARY EDUCATION

July 21, 1981
Washington, D.C.

Good day. My name is Steve Solomon. I am Co-Director of the Committee for Fair and Open Testing, a non-profit research organisation affiliated with the Center for the Study of Politics and Education, Inc. I was previously the Director of the New York Public Interest Research Group's (NYPIRG) Testing Project. My goal here today is to provide this Committee and the public with the Committee for Fair and Open Testing's views toward disclosure legilation, the Committee's concern that standardized federal legislation would be better than state statutes for the testing industry, and an up-to-date progress report on how well the law is working in New York. Lawrence Kronick and Debra Ettinger, Associates of the Committee, assisted in the writing of this testimony.

The American public's proclivity to believe in concepts adorned with numbers, the politician's need for some objective method of making our educational structure accountable and the testing industry's promotional practices all help to explain why each year over fifty million standardised multiple-choice tests are administered to America's students and workers. Indeed, the standardized testing industry has grown to become the "cradle to grave" arbitrator of social mobility and one of the most powerful corporate entities in the United States. These tests influence, among other things, whether or not one is placed in elementary school programs for the mentally retarded or for the gifted, awarded a high school diploma in over a dozen states, accepted into most American undergraduate programs, 75% of all graduate school programs, and 100% of the nations law schools. Standardized tests also play a part in determining who is given the opportunity to sell insurance in Illinois, fight fires or walk a police beat in Philadelphia, fix cars in Massachusetts, design buildings in Ohio, practice law in forty-two states, or receive a scholarship from the State of Georgia, Union Carbide, or the Brotherhood of Steam Fitters.

Yet despite the importance of these tests, the testing process has been shrouded in secrecy. Vital information concerning tests as well as test questions, have been shielded from public scrutiny by the testing industry. In 1978 the late Dr. Oscar K. Buros, winner of the Educational Testing Service (ETS) Distinguished Service Award and editor of the authoritative Mental Measurements Yearbook, lamented, "It is practically impossible for a competent test technician or test consumer to make a thorough appraisal of the construction, validation, and use of standardized tests.....because of the limited amount of trustworthy information supplied by the test publishers.

With the enaction of New York's "Truth-in-Testing" law on July 13, 1979, the secrecy surrounding the testing process has begun a slow metamorphisis. New York's "Truth-in-Testing" law, which the Federal bill is modeled after, is primarily disclosure legislation, requiring the testing industry to provide students and researchers with some basic information about how standardized tests are constructed and scored. Specifically, the law mandates that the publishers of college and professional school tests:

- Provide test-takers, upon request and at cost, with a copy of the questions scored, their individual answer sheets and the correct answers;
- File with the Commissioner of Education a copy of the test questions, the acceptable answers and all studies relating to their tests' reliability, validity and possible cultural bias; and,
- Inform test takers of the purpose of the test, policies concerning scoring errors, length of time that test scores are kept on file and the relationship between test scores and predicted performance.

WHY IS A PREEMPTIVE FEDERAL BILL NECESSARY ?

Despite the claims of the testing industry, "Truth-in-Testing" in New York State and on the Federal level is not regulation. The bill is simple information legislation, similar in concept to Truth-in-Labeling and Truth-in-Packaging laws. In no way does it regulate the content, administration or development of these tests, it only requires that the testing process be open to the public.

Now that two states have enacted their own Truth-in-Testing measures and twenty-four other states are considering similar measures, it makes sense to have a single preemptive national statute. The Law School Admissions Council and the American College Testing Program have testified at the state level that a federal bill would be preferable to having more states enact their own Truth-in-Testing laws. Bruce Zimmer, Executive Director of the LSAC, in his testimony before the New Jersey State Legislature on February 11, 1981, asked the following rhetorical questions referring to the difficulties the testing companies face if a variety of schemes of disclosure are enacted:

> What will happen if a state enacts as law the provision introduced in Texas in 1979 that even equating and pretest questions be disclosed? Or if conflicting time deadlines and notice requirements are introduced in Massacheestts, Maryland, New York, and Connecticut? What are our member institutions to do if one state protects user institution privacy and another does not? What shall we do if a state enacts a provision that provides some benefit to local test candidates "without any fee or charge?" Shall we recoup our costs only outside of that state? Or shall we add these costs to our basic fees for everyone across the nation automatically?

Clearly Mr. Zimmer has a legitimate concern. There is a great need for a preemptive federal statute that will standardize the disclosure process making compliance reasonable, practical, beneficial and economical for all parties concerned. In addition, it seems logical that due to the traditionally interstate nature of post-secondary education that any law dealing with this process should be enacted by the federal government.

Last year, before this committee the testing industry successfully argued that before any federal action be taken, the ramifications of the law's enaction in New York must be assessed:

1. The testing companies claimed that test quality would be threatened if they were required to release the questions for public inspection. Truth-in-Testing proponents predicted an improvement in test quality. Now that two years have elapsed since the New York law was passed, many testing companies and their employees have acknowledged that test quality has improved due to disclosure. Students have discovered errors affecting over half a million test-takers as a result of New York State's "Truth-in-Testing" law, many of them being faulty questions. Bruce Zimmer, of the LSAC, explained:

   > "Test disclosure promotes confidence in the integrity of the test instrument and the testing process. It is symbolic of our public accountability. Indeed it helps us in our quality control efforts."

2. The testing industry claimed that the law gave students rights they didn't want and wouldn't use. Truth-in-Testing proponents stated that students were very interested in their performance on these tests and would take advantage of this opportunity if properly informed. On the May 1980 Law School Admissions Test (LSAT) over 69% of the test takers requested a copy of their questions and answer sheets. As a result, the LSAC decided that next year LSAT test takers will automatically receive a copy of their test questions and answer sheets (see section on How the Bill is Working in New York).

Not only is the New York law working well, it has, in conjunction with the possibility of other states and congress enacting similar statutes and the discovery of so many errors, caused many of the testing companies to extend nationwide some of the benifits the law provides to New York test takers. Despite this seemingly altruistic change in policy, test publishers are still opposed to Truth-in-Testing in concept and have only implemented their national disclosure policies to hinder the efforts of other states and Congress to enact their own laws. Once the possibility of new laws is removed, the test companies could very easily resind their national disclosure policy. It is imperative that statutory guarantees be provided so that test takers now and in the future can be sure that the test which they have taken is scored correctly, reported correctly, and contains no faulty questions.

AMMENDMENTS TO THE BILL

In addition, federal legislation can be used to improve upon the New York law. New York is working well, but more needs to be done. The following are several amendments that should be added to the bill. They require the test publishers to provide more comprehensive information, which will increase public understanding of the testing process:

1. Test publishers should be required to report their scores in a 95% "true score range " (i.e., the range of scores that 95% of the test takers who take the test once and then retake the test without taking a coaching course or learning of any kind would receive. For most ETS tests this is a 120 to 140 point range -- a 500 would be reported a 440 - 560).

2. The test publishers would have to provide the Department of Education, for public review, a description of every complaint they have received from students who believe that there was an error made on their examination.

3. That if a test publisher incorrectly scores, reports or administers a student's test, the test taker should receive his/her money back plus the right to re-take the test at no extra cost within 40 days of the error being discovered.

THE NEW YORK EXPERIENCE

The New York State legislature enacted Truth-in-Testing because New York Congress of Parents and Teachers (PTA), New York Educators Association, New York Public Interest Research Group (NYPIRG), New York Personnel and Guidance Association, National Association for the Advancement of Colored People (NAACP), Puerto Rican Legal Education and Defense Fund, New York State Consumer Protection Board, New York State English Council, National Conference of Black Lawyers, New York's Attorney General's Office, dozens of research scientists, including three former ETS employees, joined every major New York student group in enthusiastically supporting the measure.

These individuals and organizations supported the measure because they were:

* Concerned about test publishers' claims that they did not make mistakes, when nearly everyone seemed to know someone who had trouble with ETS processing or scoring their tests. Documentation of fourteen separate errors, affecting 200,000 students' test scores, prior to the enaction of the New York law is in Appendix B (An in-depth explanation of the eight errors discovered due to the law and their effects on the testing industry is discussed in the Error section).

* Concerned about the fact that ETS and other publishers awarded students precise aptitude scores when their own statistical studies indicated that, due to test error there is a one in ten chance that a student who retakes a test will have a score which will fluctuate by over 100 points (fifty points in either direction). This fluctuation is important because on most ETS tests a 500 is in the 50th percentile, whereas a 600 is in the 84th percentile.

* Concerned that ETS has denied independent researchers access to the data necessary to evaluate their test's validity, reliability, and culture-biases (for one example of this, see the artical by Tamar Pacher in Appendix C).

* Concerned about the fact that in the one case prior to the enactment of New York's Truth-in-Testing law where ETS released a test form -- the February 1973 Multi-State Bar Exam, law professors from four faculties disagreed on over 35% of the correct answers.

* Concerned about the fact that researchers from the National Conference

of Black Lawyers reported that black law applicants who had comparable grades (± .01 GPA), identical majors, and attended the same undergraduate institutions as their white counterparts scored over 109 points lower on the LSAT.

Difference in LSAT Scores Between White and Minority Law School Applicants Who Have Comparable Grades, Majors, and Undergraduate Institutions

Minority Group	Number in Sample	Score Difference	
		Mean	Median
Blacks	722	-109.6	-101.1
Chicanos	352	- 99.4	-106.5
Native Americans	48	- 77.8	- 79.9
Asian Americans	277	- 34.5	- 23.8

* Concerned about the fact that the LSAT sample booklet features dozens of famous men -- presidents, senators, journalists, scholars and the like -- and not a single famous woman. This is important because several studies, including some by ETS, report that women score better on questions featuring females or persons of no specified sex while men score higher when males are featured. And the questions that do feature women are disturbing. For instance, this item from the LSAT makes some interesting assumptions. It reads as follows:

> 20. Many cultures do not share our assumptions that women are innately passive and emotional, that their roles must be domestic, and that they cannot function effectively in positions of authority. [1]

* Concerned about high school students who are taking courses on how to take tests instead of Algebra, Chemistry or Art. Concerned because teachers are altering their curriculum to teach the tests. Writing in the New York Times Magazine of September 2, 1979, Thomas C. Wheeler made a strong case that the objective test system has been a "devastating influence" on students' writing ability. "The American language -- supple, imaginative and alive -- has lost ground to the pretense of measurement," he wrote. And of the SAT in particular, Wheeler concluded, "Its verbal part, like other objective tests in language, does not ask for writing; and because it doesn't, the act of writing has withered in our schools."

* Concerned about the fact that in 1979, the American College Testing Program asked five health experts to eliminate the obviously incorrect answers from their 126 item multiple-choice health care proficiency test. ACT researchers planned to estimate the difficulty of their test by the number of answers that remained. To ACT's chagrin, each expert eliminated, as being obviously incorrect, between 9 and 26 of the answers ACT stated were correct.

* Concerned about the fact that ETS deceminated false statements concerning coaching's impact on test scores and on the costs involved in implementing Truth-in-Testing legislation.

[1] See Appendix for further test questions.

* And finally, concerned because one number from a test publisher can unalterably affect a person's life; yet that person must blindly trust the testing companies that the test is fair and is scored correctly.

The Educational Testing Service (ETS) and the College Entrance Examination Board (CEEB), the bill's major opponents, reacted to the legislation by hiring the lobbying firm that represented the State's tobacco and pharmaceutical interests. Initially, these two "non-profit" firms and their lobbyists argued that testing is a very complex technical issue that could not be dealt with in the legislative arena. In addition, they argued that there were few, if any, problems with testing, thus legislation was unnecessary.

It soon became apparent that this simplistic and condescending strategy was not convincing legislators, and ETS and CEEB committed the worst sin a lobby can committ -- they lied. A few illustrations of ETS and CEEB's intentional distortion of facts concerning the consequences of New York's Truth-in-Testing bill are as follows:

* An ETS "Memorandum-in-Opposition" circulated to New York legislators and educators in May and June 1979, stated that the New York Truth-in-Testing bill would require public filing of "all statistical data and reports" filed by the testing companies including "confidential information .... about test subjects" when in fact the bill mandated that no information identifiable with the test subjects be disclosed.

* ETS and CEEB mailgrams and memos distributed to every high school principal and college president in New York State which claimed the bill "would seriously disrupt equating procedures," when in fact questions used solely for equating were specifically exempted from the bill (equating is the process for making tests given at different times comparable to each other).

* On June 13, 1979, the day before the New York Senate was to vote on the bill, ETS distributed a "Memorandum-in-Opposition" to all New York State legislators. The Memorandum stated that:

Publication of the test questions will increase the cost of the:

SAT: between $4.73 and $12.73
GRE: between $9.73 and $32.73

LSAT: between $21.60 and $35.60
GMAT: between $14.13 and $26.63

* ETS argued that large fee increases were necessary to offset their additional costs due to having to write new questions for every test. ETS eventually retracted the above statement when:

1. Truth-in-Testing supporters publisized two internal ETS documents which revealed that test development was only a small portion of costs (33¢ to 86¢ per test in one study, 17¢ per test in another. See charts I and III).

2. Several former ETS employees testified that approximately 70% of the items of ETS' tests were made up new for each form. ETS therefore would have to increase its item writing capacity by about 50%. Since the total test development costs for the SAT, GRE, LSAT, and GMAT were less than a dollar per candidate, Truth-in-Testing legislation would only increase students' test fees by a few cents.

3. IRS documents revealed that between 1974 and 1976 ETS and CEEB earned over $10 million in "non-profits" from just their Admissions Testing Program (see Chart II).

There are several other examples of ETS and CEEB's use of inuendo, unsupported statistics, and selective misinformation to promote their programs and mislead legislators, researchers and the general public. Two specific examples are coaching and the ETS and CEEB operated Student Search Severvice (see Appendix A).

Chart I: How Much of Candidate's Fee Goes to Test Development?

Test Program	College Board	LSAT	GRE	ATGSB (now GMAT)
Fee Paid By Candidate	$5.75	$13.50	$8.00	$10.00
Total Cost (per session candidate) to ETS	$4.82	$10.83	$7.17	$ 9.22
Cost of Test Development	$ .33	$ .49	$ .86	$ .44
Test Development as % of Total Cost to ETS	6.5 %	4.5 %	8.7 %	4.8 %

Source: "Activity Analysis," Internal ETS Study, January 31, 1972

Chart II: Admission Testing Program Revenue and Expenses

	Revenues	Expenses	Surplus	% Profit
1974-75	$16,036,276	$12,550,541	$3,485,735	27 %
1975-76	$16,260,652	$13,232,474	$3,028,178	22 %
1976-77*	$17,640,000	$14,099,000	$3,541,000	25 %

*projected

Source: College Board Statement of Revenues, Expenses and Fund Balance, Year Ended July 30, 1976, with Comparable Figures for 1975 and for 1977 Current Projections.

Chart III: ETS Test Development Spending Breakdown and Budget for College Board Admissions Testing Program (SAT and Achievement Tests)

	1976 Actual	1977 Budget
Test Construction Expenses	$ 291,000	$ 389,000
Program Administration Expenses	$10,079,000	$10,864,000
Test Construction as % of Program Administrative Expenses	2.9 %	3.6 %
Admissions Testing Program Volume (in number of Test Subjects)*	$ 1,737,000	$ 1,430,000
Total Construction Expenses per Candidate**	$ .17	$ .27

*Test Subjects Volume figures are from ETS In Fact, and ETS publication
**Expenses Per Candidate calculated by dividing Test Construction Expenses by Candidate Volume

Source: "ETS CONFIDENTIAL Project Operating Statement by Pos.- Category/Project, Period Ending 3/31/77," a 66 page computer print out

HOW IS THE BILL WORKING IN NEW YORK ?

On January 3, 1980, just three days after the law went into effect, the test publishers announced that twenty of the twenty-six tests covered by the new law would no longer be administered in New York. This occurred because the testing companies took an inflexible position against any form of the legislation and made no suggestions for amendments while the bill was being considered. This resulted in a law which was insufficiently clear to insure the most effective administration. After the New York legislature adopted several amendments in February 1980, test publishers reinstated fifteen of the previously withdrawn tests. The following is a brief synopsis of specific test company reactions:

* The AMERICAN DENTAL ASSOCIATION (ADA) initially announced that they would withdraw their tests from the state. In November 1979, after meeting with state legislators, the ADA pledged to administer the Dental Admissions Test on a normal schedule and to keep test fees the same. State legislators, in turn tentatively agreed to exempt for 1980 the "perceptual ability" section of the test from the law.

* Four days after Governer Carey signed the Truth-in-Testing bill into law, the ASSOCIATION OF AMERICAN MEDICAL COLLEGES (AAMC) announced that they would not administer their test in New York. In November 1979 the AAMC filed suit in federal court charging that the law, by forcing them to make public their test questions, violates federal copyright laws.

On January 21, 1980 U.S. District Court Judge Neal McCurn granted the AAMC a preliminary injunction until the case is heard. As a result, the Medical College Admissions Test (MCAT) is exempt from the disclosure provision of the law until the case is heard and was not withdrawn from the state. A recent issue of the Columbia Law Review (January 1981) rejects the AAMC's arguments against the legislation, and concludes that the Truth-in-Testing law should be allowed to stand. Also, New York's medical schools and the American Medical Students Association (AMSA) have told Senator LaValle, sponsor of the New York bill, that they were not consulted and do not endorse AAMC's decision to withdraw the MCAT from New York or AAMC's decision to file suit in opposition to the law.

* The PSYCHOLOGICAL CORPORATION, a subsidiary of Harcourt, Brace, Jovanovich, Inc., withdrew its tests from the state. Most of these tests have an extremely low-volume and are now exempted by amendments to the law.

* The EDUCATIONAL TESTING SERVICE (ETS) and the COLLEGE ENTRANCE EXAMINATION BOARD (CEEB) continued their efforts to undermine the law. They made it extremely difficult for students to obtain materials made available by the law. They instituted a $1.75 surcharge to all New York test takers in an effort to "punish New York," and they charged an extra $4.76 to those who requested the information. CEEB also reduced by 25% the number of test dates for the SAT and cut back by 75% the number of test dates available to students who cannot be tested on Saturday for religious reasons.

In addition, they provided little if any information to students or educators that additional materials are now available. George Hanford, President of the College Board, told the New York Times, after the bill was enacted, "Why should we go out of our way to push something we don't think is useful." SAT takers must fill out a special form to obtain a copy of the test. In several instances this form has not been included with the regular registration form.

* The reaction of the LAW SCHOOL ADMISSIONS COUNCIL (LSAC), the GRADUATE RECORDS EXAMINATION BOARD, and the GRADUATE MANAGEMENT ADMISSIONS TEST BOARD sponsors of the LSAT, GRE and the GMAT, respectively, were much more favorable. They clearly informed students of their rights and broke with ETS and the College Board, extending the benefits of New York's law to students throughout the U.S. The LSAC enabled students to receive the additional information without raising prices, initially. They, unlike ETS, enabled those individuals who wanted the materials to just check a box on the registration form.

After 65% of those taking the May 1980 administration of the LSAT requested the information provided by the law, the LSAC in an effort to compliance cost effective decided to make disclosure not only national, but mandatory. Starting next year all test-takers will receive a copy of the questions scored as well as a copy of their answers. Bruce Zimmer, Executive Director of the LSAC stated, "The type of persons who apply to Law School are quite interested in their individual rights."

Recently, ETS and other members of the testing industry have been forced to reevaluate their position with respect to compliance. Eighteen months ago ETS Vice President for Law Programs, Thomas O. White, wrote an internal memo to ETS President William Turnbull (Fall 1979 Testing Digest) stating that since Truth-in-Testing had become law, publishers should do their utmost to comply with it. Today White is President of a new non-ETS affiliated organization, the Law School Admissions Service (LSAS) of Newtown, Pensylvania, which administers several programs previously run by ETS (including the LSAT). He told The New York Times (November 23, 1980) that as a result of the New York bill: "People can see more of the elements of a process that is critical to their career decisions, and they can correct any mistakes." ETS Senior Vice President for Testing, E. Belvin Williams, told The New York Times education editor Edward Fiske that New York's Truth-in-Testing bill has prodded his company into a "healthy re-examination" of the organization's internal procedures. Williams concluded that due to Truth-in-Testing"I can say with absolute certainty that the quality did not go down on any of our tests. To the contrary, it is better than ever."

Yet despite the high request rate for the LSAT and CEEB's new disclosure policies, ETS and CEEB have spent thousands of dollars on high priced lobbyists, to prevent other state and Congress from extending the benifits of Truth-in-Testing to their constituents. In August 1978, ETS Presdent William Turnbull wrote a memorandum titled Public and Professional Attitudes Towards Testing and ETS. The 21-page confidential document outlines Turnbull's plan to spend $450,000 to influence legislators, government officials, teachers' unions and civil rights organizations (see Appendix D). Testing lobbies in other states have continued their efforts, presenting questionalble arguments against the bills. For example:

1. they claim that the New York law would be repealed, when in fact the legislature in New York never considered such action.

2. legislators in other states were told that "Truth-in-Testing" is really a federal issue; when at the same time ETS lobbyists were telling members of Congress considering the bill the the Federal government had no right to interfere in education; and,

3. the cost of administration of the law would be in the area of 50,000 dollars when in fact the New York State legislature for the past two fiscal years has budgeted absolutely nothing for the laws administration.

In February 1980, six weeks after the New York bill became law, Senator LaValle told the following story to illustrate his frustrations with ETS and CEEB officials: "In the early 1960's when Congress enacted auto emission regulations, both General Motors and Toyota hired 1,000 ne employees. Toyota hired 1,000 engineers and GM hired 1,000 new lawyers." The following are just a few of the lobbyists ETS and CEEB have hired to defeat Truth-in-Testing legislation:

* Florida - Bruce Smathers, the state's former Secretary of State;

* Maryland - Tom Lord, the state's former Assistant Attorney General;

* Massachusetts - Tom Joyce, the lobbyist retained by that state's banking and nuclear interests;

* New Jersey - Joe Katz, Lobbyist for that state's restaurant, motel, cigarette, auto dealers, race track and casino interests; and,

* Washington, D.C. - the law firm of Wilmer, Cutler and Pickering, whose other clients include General Motors, IBM and DuPont.

ERRORS AND THEIR RAMIFICATIONS

New York State's Truth-in-Testing law has resulted in increased public scrutiny of tests as well as provided college and graduate school applicants with the right, after scores are reported, to a copy of the questions asked and their answer sheets. Prior to the enaction of Truth-in-Testing in New York, students had little if any way of detecting or proving that there was any problem with test items or reported scores. In fact, as mentioned in the accompanying article, ETS claimed that they were virtually infallible. Since the enaction of the law, errors effecting the scores of nearly one-half million students have been discovered. Six different categories of errors have been brought to light: administrative errors, factually incorrect questions, scoring errors, reporting errors, ambiguous questions, and biased test items.

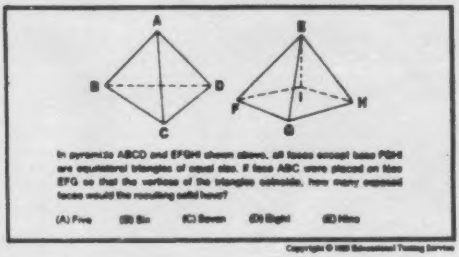
The first error detected after the passage of the law was discovered by a Pittsford, New York high school student. He wrote ETS pointing out that one of the test's reading comprehension passages that dealt with how migrating birds navigate, was incorrect. "Reading false statements does tend to jar one's senses," he said, "possibly causing errors to be made in answering the questions."

Next, Michael Simon, a recent graduate of Columbia University's Masters Program in International Affairs, who took the LSAT last summer wrote ETS that their computer had incorrectly marked a "no answer" in five places where he had entered a reply.

Similarly a graduate of Arizona University, now a New York resident, took the LSAT in February 1981. Approximately three weeks later she received a score of 564 on a scale of 200-800. Several weeks later she received a copy of the questions scored, her answer sheet and the correct answers to the questions. When computing her score from the information provided she found that her score was actually a 624, 60 points higher than the LSAS had reported. After several rechecks she phoned ETS, who acknowledged their mistake, claiming a computer had errored. ETS then issued corrected reports to the law schools she had applied to for admissions.

Without Truth-in-Testing, errors like these would never have been discovered. In April 1980 NYPIRG and the Committee, in an effort to determine the frequency of such occurences, sent questionaires to the nine major testing companies asking for information. To date they have received responses from two agencies, the College Board and the Graduate Records Examination Board. Both letters acknowledged the request and stated that further information would be forthcoming. Unfortunately NYPIRG and the Committee have received no further information.

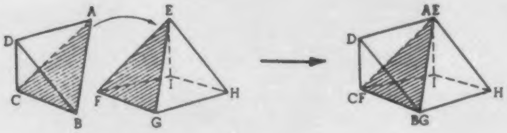
The fourth error discovered got quite a reception, making the front page of The New York Times on March 17, 1981. Danial Lowen, a 17 year old Cocoa Beach, Florida High School student, proved that the answer ETS and its panel of 16 college professors selected on a question from the October 21, 1981 PSAT was inferior to another choice. Senior Vice President for Testing Programs at ETS told The New York Times that the question was a "lousy item." Why ETS and the College Board kept this error a secret for three months still remains unknown. Could it be that this information could have had a profound effect on pending legislation?



In pyramids ABCD and EFGHI shown above, all faces except base FGHI are equilateral triangles of equal size. If face ABC were placed on face EFG so that the vertices of the triangles coincide, how many exposed faces would the resulting solid have?

(A) Five (B) Six (C) Seven (D) Eight (E) Nine

Copyright © 1980 Educational Testing Service



Depending on whether the faces considered are those of the original two solids or those of the combined solid, both choices (A) Five and (C) Seven have merit as correct answers. When the question was originally scored, only choice C was considered correct. Although seven of the original faces remain exposed, the solid that results if face ABC is placed on EFG contains but five distinguishable faces. Some of the faces of the original figures (for eample, faces ABD and EGH in the diagram on the bottom) lie in the same plane and form a parallelogram when the solids are placed together. The diagram provides an illustration.

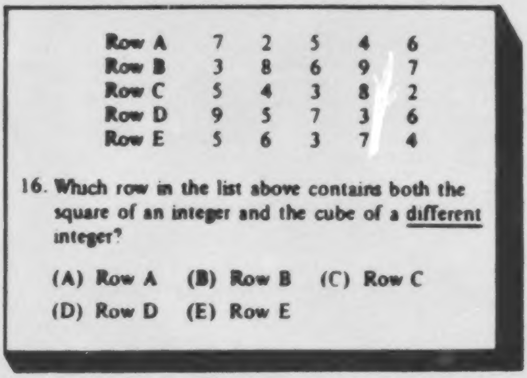
The October 1980 PSAT was the first PSAT ever released to the public. This disclosure of the PSAT was a result of a new College Board policy, introduced due to the increasingly apparent success of Truth-in-Testing in New York and the impending threat of the passage of a similar bill by one of the many other states considering this type of legislation.

Mr. Lowen's discovery resulted in 240,000 students' test scores being upgraded, 60 additional students being awarded financial scholarships and over 100 additional "letters of commendation" being awarded by the multi-million dollar National Merit Scholarship Program who annually awards students with high PSAT scores.

In addition, the test Mr. Lowen had taken was made up of old SATs. Robert Moultrop, ETS spokesman, estimated that the scores of about 80,000 of the 316,000 persons who took tests using the "lousy item" in recent years should probably be raised by as much as 10 points. Their scores have never been corrected.

Mr. Lowen's discovery, to quote a New York Times headline, opened a "Pandora's Box." The Times and the testing companies received a "deluge of letters" suggesting a variety of new answers to the same questions. E. Bevin Williams, Director of Test Development at ETS in response to queries on the legitimacy of alternative solutions presented in the correspondence said with a shrug, "It seems as if every answer was correct except ours."

Just one week later on March 23, 1981, the fifth mistake, discovered by Michael Galligan, was announced by ETS. Galligan, a senior at Clarkstown South High School in New City, New York, discovered the error on the October SAT. Again the public was not informed until well after the error was discovered and rescored. In a letter to ETS, Mr. Galligan pointed out that a question about integers on the exam had two right answers. Despite admitting that Mr. Galligan was correct, Ernest Kimmel, Director of Test Development at ETS, declared, "We know the people who chose C (the other correct answer discovered by Galligan) were less able: they skipped over the obvious answer."



Row A	7	2	5	4	6
Row B	3	8	6	9	7
Row C	5	4	3	8	2
Row D	9	5	7	3	6
Row E	5	6	3	7	4

16. **Which row in the list above contains both the square of an integer and the cube of a different integer?**

(A) Row A **(B) Row B** **(C) Row C**

(D) Row D **(E) Row E**

COPYRIGHT © 1980 BY ETS AND THE COLLEGE BOARD

A student pointed out that (C) should have been accepted as a correct answer, as well as the "official" (B).

As a result of Mr. Galligan's discovery about 16,000 of some 84,000 "less able" test takers in New York State had their scores increased and several hundred thus qualified for state scholarships worth $1,000 over a four year period.

Mr. Lowen's discovery, coupled with Mr. Galigan's caused quite a disturbance in the testing industry. Just three days later, the College Board Trustees, faced with convincing evidence that opening up the testing process had and would improve test quality, not to mention the effect these discoveries would have on pending legislation, voted to partially extend the "Truth-in-Testing" concept nationwide. Their new policy, however, is limited to selected test dates and unlike "Truth-in-Testing" legislation, provides no concrete guarentees that disclosure will continue in the future. It is interesting to note that College Board officials just three days earlier testified against Truth-in-Testing in both Massachusetts and Florida.

Many things concerning the sixth and seventh mistakes which occurred on the GRE in advanced engineering and advanced psychology are still unexplained. Both questions, according to ETS, were not released by the GRE Board because of the possibility of the questions being reworded and used in future administrations of these exams. Who discovered the errors is also open to debate. Edward Fiske, New York Times Reporter, on April 28, 1981 said that, "students successfully challanged questions on the GRE." ETS Public Relations person, Marilyn Ballis claimed that both mistakes were discovered as part of the ongoing review process after they were printed and administered. Assuming Mr. Fiske was correct, students who had discovered these mistakes would have gained no satisfaction, for the GRE Board voted not to change any incorrect scores unless differences of more than 10 points would result.

Ms. Ballis, in referene to the two errors remarked, "it was only a matter of four points" and " was not statistically significant." Robert Altman, acting Vice-President for Higher Education Programs at ETS, in defense of the Boards decision stated, "We continually tell colleges that it is dangerous to make judgements about students on the basis of 10 point differences. If we sent out corrections of this magnitude, it would contradict everything we have been telling them." When asked about the GRE Board's decision, Thomas White, President of the LSAS, remarked, "I wouldn't want to apply the same logic to our test or to the SAT. There are instances, even if we do not consider them justified, in which 10 points can have an impact on a person's educational opportunities."

A recent FTC report entitled "Coaching for Standardized Admissions Tests" reprted that "42% of the private four year colleges and 37.9% of the public four year colleges responding (to their survey) indicated that they had minimum SAT score requirements below which applicants generally are not considered eligible for admission." In addition, although test makers "discourage" the use of cut-off scores, preliminary selection for National Merit Scholarships are based on "cut-off" scores on the PSAT. Finally, according to Bevin Williams, the testing industry isn't always aware of how test scores are used. He stated, "As long as there is any chance that not rescoring will affect any individual, then it is irresponsible not to do so."

Of the 23,200 students who took the February 21 administration of the LSAT, almost 22,000 had to have their scores changed. The eighth error was discovered by Christopher Laucks of Brookline, Massachusetts when he realized that their were two possible answers to a Geometry question on the exam, (see diagram). The discovery resulted in 15,000 students' scores incresing, 1,600 scores remaining the same, and 6,600 scores dicreasing.

Figure 1
Question 23 as it appeared on the February 21, 1981 LSAT

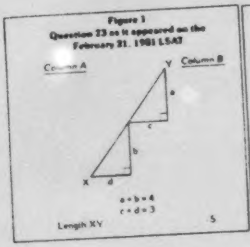
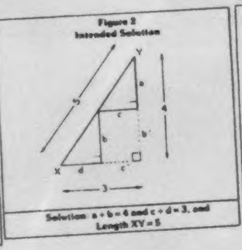
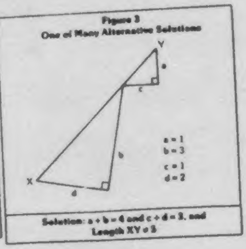


Figure 2
Intended Solution



Solution: a + b = 4 and c + d = 3, and Length XY = 5

Figure 3
One of Many Alternative Solutions



Solution: a + b = 4 and c + d = 3, and Length XY ≠ 5

EXPLANATION

Figure 1 shows question 23 as it appeared on the LSAT This question involves the comparison of two quantities, the length of XY and the integer 5 The credited response to the question was (C), that is, that the two quantities are equal The intent of the question was to have the candidate observe that, if line segments labeled a and d were extended, they would meet in a right angle and form a right triangle with hypotenuse of length XY and sides of length a + b = 4 and c + d = 3 (see Figure 2) The length XY would then be $\sqrt{4^2+3^2}=\sqrt{25}=5$, and the answer would be (C)

This reasoning is correct if, as many people assumed from the way the figure is drawn, the angle between the two triangles where line segments labeled b and c meet is a right angle or if line segments labeled a and b (or those labeled c and d) are known to be parallel However, since this information is missing, there can be other triangles that satisfy the given conditions, but for which the length XY is not 5 The person who made the initial inquiry pointed out one such case (see Figure 3) In this case line segments labeled c and d are not parallel, and the length XY is $\sqrt{3^2+2^2}+\sqrt{1^2+1^2}=\sqrt{13}+\sqrt{2}$, which is greater than 5 Since there are cases for which the length XY is 5 and other cases for which the length is not 5, the answer to the question as it appeared is (D), that is, the relationship cannot be determined from the information given.

Three days after the mistake was reported, the LSAC, which represents 171 law schools, approved by the American Bar Association and is sponsor of the LSAT, announced drastic changes in the structure and scoring of the LSAT. According to the Associated Press, begining in the summer of 1982 students taking the LSAT would no longer be required to correct ungramatical sentences or solve geometry and other "quantitative" math questions. Also the scal used to report LSAT scores (200-800) would be replaced by a scale ranging from 10-50. Bruce Zimmer explained the council's actions are an effort to "discourage the use of small score differences to shape admissions decisions." It is important to note that generally Law Schools plac 60 to 70 per cent of their emphasis in admissions on LSAT scores. Only 30 to 40 per cent is placed on other criteria, like GPA and extra-curricular activities.

DEMISTIFICATION

Truth-in-Testing legislation in New York is also having a positive impact on the availability of quality test preparation courses to low income and minority test takers. Patrick Shields of the East Harlem College and Career Counseling Program, a federally funded organization designed to provide additional assistance to college and career bound students, has been providing a three week coaching course for program participants. He characterized the materials available prior to the enaction of the law (published commercial materials and ETS sample booklets) as of questionable value. There was a very limited number of sample items of which many were of perhaps little value.

Due to the enaction of the law, Mr Shield maintains that having realistic SAT items available has enabled him to do a "better job." "Students," he described, "feel more comfortable about taking the tests now that they've had their hands on the real thing."

Jeannette Burgos-Servedio, Head Counselor of Talant Search, also feels that Truth-in-Testing is providing information to her which helps her to do a better job. In a letter to the Committee she states:

> At the request of students in our program, we at Talent Search Program of Bronx Community College sponsor an SAT coaching course ever academic year. Presently, we charge the students $10 for an average of eight two-hour sessions. We are now in the process of writing the course into our project so that we an offer the course at no cost and, therefore, make the course available to an even greater number of our student population.
>
> Throughout the 1980-81 academic year, we had four sessions; and we got a opportunity to use actual SAT tests with the students in the last session. Their responses, in comparison to those students who only got an opportunity to use sample tests, were more positive. They felt that having had an actual test before them helped not only prepare them more adequately for the exam, but they also felt that the level of familiarity they gained by having the test would help relieve much of the anxiety they would have gone into the exam with. This, in turn, would enhance their chance of obtaining a better score.
>
> As a counselor, I feel that, regardless of what others may feel about coaching, coaching does help our student population. That population being predominantly low income, minority students. Coaching for these students gives them an opportunity to learn test taking techniques they do not necessarily pick up in the high schools. It also raises their level of confidence.

According to Luis DeGraffe, of the Puerto Rican Legal Education and Defense Fund which administers a coaching course to about 60 paw school candidates every year, Truth-in-Testing helps his students in two major ways:

1. students who take the exam more than once are enabled to see what part of the exam they are weakest in. They can then prepare accordingly.

2. it enables us to institute a simulation of a test situation. This can help to relieve much of the test anxiety which students generally experience.

The ultimate solution to the problem of disclosure is for the federal government to enact a nation-wide "Truth-in-Testing" law. This will eliminate the difficulties the testing industry faces having to comply with a variety of forms of state legislation and the costs incurred thereof. In addition, students now and in the future will be assurred that the testing process is conducted in a fair and open manner.

APPENDIX A

Coaching

In late 1976, the Boston Regional Office of the Federal Trade Commission (FTC) began an industry-wide investigation into the validity of claims made regarding the effectiveness of coaching for standardized admissions tests. They subpoened the files from all of the commercial coaching schools, as well as from the testing companies. Ironically, although their "investigation was spawned by concern that commercial coaching schools were misleading prospective students by deceptive claims," they found the opposite to be true. In fact, they reported that students who took a coaching course before taking the SAT a second time, score 100 points higher than those students who simply retook the SAT without taking a course. Additionally, research, including some done by ETS, has shown that coaching improves Minority students' scores the most -- yst minority students are probably the least able to attend coaching schools, which cost as much as $350.

In the Spring 1980 Harvard Educational Review, Drs. Warner Slack and Douglas Portor of Harvard Medical School report in "The Scholastic Aptitude Test: A Critical Appraisal" that

> In publications sponsored by ETS or the College Board, wee found nine cinclusions that studies to date had demonstrated coaching to be ineffective, even though the published experimental evidence indicated the contrary. We also found seven assertions that gains in scores from coaching were no significant, although they were statistically significant.

The Harvard suthors concluded that the:

> Educational Testing Service's technical and promotional publications are characterized by omissions and misleading statements. They have enhanced the product's image at the risk of the students' self-esteem. The students' trust has been broken.

In addition, Slack and portor came across a 1968 ETS/CEEB booklet titled "The Effects of Coaching on Scholastic Aptitude Scores." This booklet summarized seven studies of coaching programs which failed to produce significant score gains. It conveniently did not mention a College Board sponsored study at the U.S. Military Academy which showed a coaching gain of 136 points on the SAT. The studies which the booklet did choose to report invloved coaching programs substantially shorter and lessintense than those offered by the commercial ciaching schools. The Harvard researchers stated that, "there is ample evidence that students can successfully train for the SAT" and that such evidence was available at the time the College Board booklet was published in 1968.

In 1972, ETS researchers Dr. Louis Pike and Dr. Franklin Evans analyzed the studies cited by the booklet on coaching for the SAT math section. They concluded that ETS' unequivocal denunciation of coaching was "hazardous."

Yet ETS continued to distribute this pamphlet until last year.

Despite convincing evidence of the effectiveness of coaching courses, the Federal Trade Commission report contended that the College Board and ETS "continued to indicate that special preparation was unlikely to be of significant benefit."

Taking the SAT, a student guide published by the College Board, made the following statement:

> The verbal and mathematical abilities measured by the SAT develop over years of study and practice. Drilling or last-minute cramming probably will not do much to prepare you for the test. However, if you are not taking a mathematics course, a review of mathematics concepts, such as those given in this booklet, will be useful.

"The Student Bulletin 1979-1980" addressed the coaching question in three sentences:

> Before the test date, read the appropriate descriptive booklet, Taking the SAT or About the Achievement Tests, which will give you a better understanding of the test. Studies have shown that cramming does not raise scores, but knowing what to expect can be helpful. A good night's sleep and a nourishing breakfast before you take a test are also helpful."

The FTC study concluded that the 1979-1980 SAT materials "did not appear to fully apprise students or educators of the possible benefits of coaching."

Student Search Service (SSS)

In 1979 ETS and CEEB's most profitable activity was the CEEB owned, ETS operated Student Search Service (SSS). Last year the SSS sold mailing lists containing over 22 million names and addresses to 1800 clients. SSS collects its lists from the forms the 1.8 million high school students fill out when they take any one of the fourteen CEEB sponsored, ETS developed multiple-choice college entrance "aptitude" and "achievement" tests. ETS' 1979-1980 Scholastic Aptitude Test (SAT) promotional booklet informs students that the:

> College Board's Student Search Service is an information service for students, colleges and governmental scholarship programs. It is free (their emphasis).... and works this way:
>
> If you ask to participate, college and scholarship agencies interested in students with your characteristics can ask for and receive your name, address, sex, high school and intended major.

Somehow ETS and CEEB forgot to inform high school students (or their parents) that:

* the SSS considered the Defense Department to be a governmental scholarship agency.

- the Newspaper Fund, Inc., a foundation affiliated with The Dow Jones, Inc., the owners of the Wall Street Journal, was also considered to be a governmental scholarship agency.

- besides the characteristics cited above, the SSS also provided its clients with information concerning the students' race, socio-economic status, and test scores.

- their names were going to be sold at a large profit. In FY 1979-80, the SSS sold lists containing over 22 million names, earned ETS/CEEB over $2.7 million in revenue and $2.1 million in "non-profits" (excess income over expenditures).

In late 1980 ETS/CEEB, reacting to legal action brought by the Committee in a Brooklyn, New York federal court, started informing students, for the first time, that they will be categorized by ethnic background and family income. Today's ETS/CEEB solicitation, however, does not inform students that their names will be provided to the military or that the non-profit corporations sell (at a large profit) the information they provide (see Testing Digest, Winter 1981, pages 1-3 in Appendix).

Appendix B - Mistakes

In 1974, Jean Kerr, ETS director of Testing services, stated:

> I think that in terms of our reporting and our processing, we have been virtually one hundred per cent correct.... I would say over the past few years, at least in my association with this division of operations, and that's really what I speak to, there have been virtually no errors."*

In the following year, Bernard Tehorni, ETS Vice President in charge of operations, was asked, "What percentage of the reports that you send out are free of error?" He responded, "Free of error? One hundred per cent."*

In November 1976 ETS Executive Vice President Robert Solomon, testifying before the Privacy Commission was "asked specifically and repeated about errors that might be correctable if students had access to their ETS files." He responded that to the best of our experiences, ETS doesn't make mistakes so no corrections would need to be made.

What follows are brief descriptions of over a dozen scoring errors and reporting errors effecting nearly a quarter million students that have occurred since 1972. These descriptions were obtained from press clippings people have mailed to The Committee during the past three years. In all liklihood a more systematic approach to discovering errors would have uncovered mistakes affecting thousands more students opportunities for graduate and professional school.

What disturbs me about the following mistakes is not that they occurred--for mistakes will always happen, but test publishers seeming inability to rectify mistakes once they've been made. Truth-in-Testing will help remedy this situation by giving students access to the information they need to verify that their tests were scored correctly.

Error 1

In January 1973, the Educational Testing Service (ETS) informed Harvey Block that he had scored a 270 out of a possible 800 on the Graduate Records Examination (GRE) in Advanced Psychology, taken in December of 1972. Harvey Block, then a senior at Yeshiva University, majoring in Psychology, with a 3.92 grade point average, was, to put it mildly shocked by his low score. He immediately called ETS requesting that they rescore his exam. ETS initially responded that they did not make mistakes or rescore exams and that many people thought they would receive higher scores than they actually achieved. Mr. Block was persistent and in late January 1973, ETS agreed that within a week they would rescore his exam. The following week Mr. Block contacted ETS only to find that no rescore had occurred and again received a similar assurance from ETS that a rescore would soon take place. Mr. Block taking no chances followed this up with a written request.

Some time between February 14 and February 23, 1973, ETS upon rescore found that an error had occurred and Mr. Block's correct score was a 570, three hundred points higher than what was initially reported.

---

*See The Reign of ETS by Allan Nairn & Ralph Nader

147

In mid-March 1973, Block received notification of the score revision. Included was a list of schools to whom corrected scores were sent. Block discovered that the revised score had not been sent to over half the schools he had applied to. During this period Block received several rejections letters from many of the schools that had not received his corrected score.

On December 13, 1974, Mr. Block, who is currently a doctoral student in New York University's Social Psychology Program, filed a lawsuit against ETS seeking damages for income lost due to his inability to enter a doctoral program. On November 17, 1979 Block lost his case. ETS admitted mis-scoring Block's test, but successfully argued that he might not have been admitted to the graduate program even if his scores had been correctly reported. During the trial ETS Director of Higher Education Programs, Robert Altman stated that Block's answer sheet was mis-scored because Block gridded the second letter of his last name incorrectly on his answer sheet as "i" instead of "l" and provided an incorrect registration number. In the process of making the two corrections an ETS clerk also revised the number indicating the specific form of the test taken by Block (which was entered correctly) to an incorrect number. Altman continued that:

> "Although we do not know why this action was taken, it caused an incorrect scoring key to be applied to his answer sheet, with the resulting incorrect score. The quality control procedures described did not uncover this error, probably because of the low incidence of similar errors at this administration date, none of which had been identified as quality control cases."

Yet Altman later went on to state:

> "In addition to the mis-grading of Mr. Block's answer sheet, eleven other answer sheets for the Advanced Test in Psychology administered on December 11, 1972 were mis-graded."

When questioned by Block's attorney in 1975, ETS claimed that they could no longer identify or produce information about the other eleven scoring errors or whether these test-takers had been notified of the error.

Block can at least be thankful that ETS discovered his mistake. What happens when the test publisher doesn't find the mistake. Equally disturbing is what happens to a student who skips a question on a multiple-choice test, but doesn't skip the corresponding space on his/her answer sheet. Such clerical errors are all to easy to make on speeded multiple-choice tests, especially when a student is under pressure.

Truth-in-Testing legislation will grant Connecticut test takers the same rights New York students now enjoy - the ability to make sure that their test was properly scored and that they filled in their answer sheet correctly.

Trenton Times, N.J., 11/16/79, Superior Court of New Jersey Law Division - Mercer County, Docket No. L-13177-74 (Civil Action Interrogatories)

Error 2

ETS' 1975 and 1976 Law School Admissions Test (LSAT) registration form asked students whether they have taken the test before. This question appeared among a series of general clerical questions (name, address, and so on). Applicants had no way of knowing that failure to indicate that they had taken the test before (whether by answering the question incorrectly or merely leaving it blank) was important.

The reason for asking this question was not clerical. ETS researchers had found out that students who retook a test, improved their scores by an average of 30 points. The increase occurred because these test-takers were now more familiar with the test's format. It should be noted that at this time ETS official position was that coaching would not improve students test scores.

According to the February 5, 1977 New Republic:

> The test score report form sent to law schools contained-unbeknownst to the student who was paying to have this information sent - a space labeled "unacknowledged repeater.". When this space was marked by an asterisk, it indicated that a candidate applying to take the test for the second time had not indicated that this was the case. The asterisk appeared even if a candidate (possibly by accident) failed to answer the question at all. Ralph Smith, a professor and admissions committee member at the University of Pennsylvania Law School, explained that the "unacknowledged repeater" designation raised "a question of (the applicant's) integrity." A number of other law school admission officers confirmed that this was the construction admissions committees put on the asterisk."
>
> Smith and several others think the appearance of the asterisk kept some applicants out of law school. Of course there are such a variety of reasons an admissions committee can give for denying admission to any one applicant that it is all but impossible to prove that anyone was excluded because of being designated an "unacknowledged repeater."
>
> In the 1975-76 applicant year, the dean of admissions at the University of Pennsylvania Law School noticed that more and more of these asterisks were appearing on LSAT score reports. He reported this to ETS and asked whether the designation was correct in all these cases. Shortly thereafter ETS responded with a memo sent to all law school admissions deans, which read in part:
>
>> It has recently become apparent that a problem exists with regard to the flagging of candidates as "unacknowledged repeaters" (an asterisk in the area under "Unacknowledged repeater" in the

> summary section of reports). This problem has been caused by several factors, some of which raise policy questions which will be considered next spring. The result for the current year, however is that some candidates who are not "unacknowledged repeaters" have been designated as such.

The memo warned that the designation should be disregarded and explained that "it will not be possible to identify all such candidates" (those wrongly identified as "unacknowledged repeaters").

There are several disturbing things about this incident. First of all, since applicants have had no way of knowing that the "unacknowledged repeaters" designation existed, discovery of this error depended on the fortuitous curiosity of one admissions dean. Second, since applicants are not permitted to see the reports ETS sends to law schools, individuals would have had no way of clearing themselves of false charges even if they had known that the problem existed. Third, even after the error was discovered, ETS kept it quiet and apparently made no attempt either to correct its past mistakes or to inform the applicants who were harmed by them. In short, candidates who were perfectly acceptable may have been denied access to their chosen profession without ever suspecting the reason.

ETS avoided any mention of the "unacknowledged repeater" incident in testimony at a Privacy Commission hearing in November. Robert Solomon, vice president of ETS, was asked specifically and repeatedly about errors that might be correctable if students had access to their ETS files, but he failed to mention the asterisk episode and insisted that ETS errors were "to the best of our experience" no problem.

Kim Masters, "ETS's Star Chamber," New Republic, February 5, 1977

Error 3

ETS changed the content of the Law School Admission Test (LSAT) between the July and October 1977 administrations, apparently without notifying the nation's law schools. The result was that nearly twice as many students scored 750 or better in October than July, with triple the number of perfect 800 scores reported. Law schools with rolling admissions policies (that is, admitting applicants when the application is completed as opposed to waiting for a set date to look at all the applications at the same time) admitted students on the false assumption that LSAT scores from October 1977 meant the same thing as those from July 1977 and earlier. Students who took the LSAT in July 1977 or earlier were, in effect, penalized in terms of their chances for admission. The score variances weren't explained to law schools by ETS until February 1978.

Comey Winfrey, "Revised Law School Test Leads to Dispute," New York Times 4/28/78, p. A23.

Error 4

> In May 1978, the American College Testing Program (ACT), which administers the Medical College Admissions Test (MCAT), wrote to medical schools to inform them that the MCAT scores they had been sent were incorrect. Due to an error in the scoring key sent to ACT by the MCAT development agency, the scores of roughly 90 percent of the test-takers were too low.

Maxwell Gould, "MCAT Scores Fall from Past Level, Prompt Regrading," Harvard Crimson, 5/19/78, p.1.

Error 5

> ETS misscored over 50,000 Graduate Management Admissions Tests (GMAT) taken in November 1977. Before ETS corrected the mistake numerous business schools had mailed out letters of acceptance and rejection.

J. Wyatt Emmerick, "Error in GMAT Scores Affect Applications," Harvard Crimson, January 18, 1978, p. 1.

Error 6

> Several South Carolina school districts require teachers to pass ETS' National Teachers Examination (NTE) as part of the job certification process. In 1977 a South Carolina teacher was fired because ETS incorrectly reported that her NTE test score was below the school district's passing rate. ETS corrected its error in 1979, but the teacher was out of work for eighteen months due to the mistake.

Spring 1980, Testing Digest, p. 2

Error 7,8,9

Kenneth Drexler, the President of the Student Government Association of North High School in Great Neck, New York, explained to the U.S. House Committee on Education and Labor three mistakes that occurred in his school in 1978:

> Our school has gotten involved in the question of the testing industry and the truth in testing bill primarily as a result of several incidents which occurred at out schools. There were errors made by the Educational Testing Service. I described them in my testimony and have given them titles. They are not official titles but are used to give identification to the incidents.
>
> First, the case of the missing paper and the duplicated test.
>
> In 1976 the Great Neck school district notified ETS that a test booklet of a SAT exam was missing. It subsequently developed that the missing booklet was taken by a Great Neck teacher who made copies and used it to give students practice in taking the SAT tests.

There was no question but that what the teacher did was illegal, but the test was used for purposes of coaching. Other examples are reported as missing and such tests should never be repeated. The teacher never contemplated, even as a remote possibility, that ETS would repeat the same test.

When ETS unaccountably did repeat exactly the same test only 2 years later, almost 50 Great Neck students, fewer than 5 of whom had even seen the earlier test, had their scores invalidated. ETS promptly brought a suit of copyright infringement against the teacher which was subsequently settled out of court.

No justice, however, was available to the 50 Great Neck students whose tests were invalidated due not to any fault of their own but to the gross inefficiency of ETS in repeating a test when a copy had been reported missing.

The October weekend fiasco: ETS, to its credit, is aware that some students observe Saturday as their Sabbath, and it gives such students the option of taking its exams on Sunday.

On Sunday, October 29, 1978, 13 Great Neck North students took the PSAT exams. Incredibly, ETS administered to them exactly the same test that was given to their classmates on Saturday, October 28.

The results: (1) The National Merit Scholarship Corp. disqualified the results of the 13 students who took the Sunday exam, and (2) to make amends, ETS asked the NMSC, which bases it scholarship awards on a formula based upon PSAT scorers, to use the regular SAT scores which the 13 students obtained the following spring in making the scholarship awards.

This has caused protests from others who claim that the scores on the regular SAT's are generally far better than those on the PSAT's, that the two are not comparable, so that in the words of a school newspaper editorial, "It's like comparing apples and oranges, or more appropriately, it's like comparing responsible organizations to irresponsible profitmaking groups such as the ETS. It just cannot be done."

How to demolish a student's college prospects: ETS permits a student to cancel an exam by filling out a cancellation report when he turns in his test paper. If he does so, ETS promises that it will not grade the exam nor make any report thereof.

One of the better students who had been rejected by some colleges later discovered that ETS had reported to all colleges to which he applied the results of an exam which he had canceled after completing fewer than half the questions.

There is no way to ascertain how substantially this ETS blunder

> affected the decisions of the colleges from which he was rejected. But how do you make restitution to this student? One must surely speculate as to how many times this has happened and gone undiscovered.
>
> All three of these events occurred in one school with a population of 1,200 students in 1 year's time.

Truth-in-Testing Act of 1979, The Educational Testing Act of 1979, Hearings before the Subcommittee on Elementary, Secondary and Vocational Education of the Committee on Education and Labor House of Representatives. 96 Congress, October 11, 1979.

Error 10

In 1979, National Evaluation Systems of Amherst, Massachusetts, made a reporting error on the scores of New Jersey's ninth grade reading test. The error, .1 of a percent, made the difference in 1,300 students failing or passing the test. The company notified school districts, who then had to notify students and parents. Before this process was completed, many students had already been told they had failed the test.

(*Testing Digest*, Winter 1981, p. 15)

Error 11

In February 1979, a graduate of the Georgetown Law Center took the Maryland bar examination. Part of this examination consisted of the Multistate Bar Examination (MBE), a standardized test administered by ETS. In April 1979 the student was told that he had failed the bar exam and would have to take it again if he wanted to pass. A few months later ETS informed the student that when someone was conducting a random quality control check, they discovered a "crinkle" in his answer sheet. This had covered part of the answers the student had made, and the computer had scored that portion as all incorrect. When the test was rescored by hand ETS found that the student had actually passed the test. ETS' Director of Public Relations told *Newsday* that "to his knowledge (this) was the first time a mechanical error (had) interfered with the accuracy of computer corrections.

"In Short: Wrinkled by a Crinkle", *Newsday*, July 31, 1979, part II, p. 2

Error 12

The Washington Post reported that ETS lost the answer sheets of 163 Mongomery County, Maryland high school seniors.

Spring, 1980 *Testing Digest*, p. 14)

Error 13

> "A leaky watermain" according to an ETS spokesperson, caused a computer to fail to send thousands of students admissions tickets for the October 13, 1979 Law School Admissions Test (LSAT).
>
> ETS' switchboards were unable to handle the thousands of calls

from students whose checks had been cancelled by ETS, but were not mailed test registration forms. One frustrated young man reported that he called ETS thirty times a day for a week without getting through.

Those lucky ones who did get through received a variety of responses. Some received mailgrams confirming their registration. Others were mailed a letter indicating which test site they should go to. Others were assured that a test would be waiting for them at their originally designated test center. Still others received two or three different responses from ETS. One student received four identical admissions letters.

Three hundred students directed to Georgetown University in Washington, D.C. found that there were only two hundred and fifty tests.

The test administrators admitted students on a priority basis: those with regular registration tickets were allowed in first; those students holding mailgrams, second; next were those with letters. Fifty students with verbal assurances found themselves out in the cold.

One such student, Steve Pearlman of Georgetown University, had completed an expensive preparation course. The course was designed to prepare him for October 13. "I took half a week off from school in order to peak on that day," he lamented.

Theresa Koncick wrote that the entire situation was "chaotic". "We were all nervous about the test and then they announced that some of us wouldn't get to take it. Since I arrived early, I got in. But the test started an hour and a half late and I'm sure the waiting around lessened my concentration and my score."

Konick continued that "on November 13, 1979 I recieved an admissions ticket" for the October 1979 test. ETS as of today has not offered Theresa Koncick or others affected by this mistake an apology or even a written explanation of what happened.

Spring 1980 Testing Digest, p. 2, 15

Error 14

In 1980 thousands of law school applicants' scores arrived late due to a computer foul-up at the Law School Data Assembly Service (LSDAS) in Newtown, Pennsylvania. The LSDAS, an ETS subsidiary, compiles law school applicants' grades and Law School Admissions Test (LSAT) scores and sends them to the nation's

220 law schools.

ETS did not publicize their mistake until New York State Assemblyman Mark Allan Siegel, Chairman of the Assembly Higher Education Committee, conducted an investigation. Siegel found that the ETS mistake had caused processing delays of about 80,000 law school applicants and that ETS had not informed students that there would be an eight week delay.

Siegel's investigation also uncovered the fact that 30,000 yellow matching forms were not mailed to students. Applicants fill out these yellow forms and send them directly to law schools. The schools then forward the forms to LSDAS. In turn LSDAS sends back the yellow forms to students to verify the arrival of their transcripts.

After learning about the legislators' investigation, ETS on January 28, 1980 mailed a press release to college newspapers announcing the eight week processing delay.

Assemblyman Siegel in a February 21 interview described the delay as "unprecedented-in terms of a complete collapse of service." He continued that ETS, by not individually informing students of the delay "had failed in their judiciary duties."

The forms the LSDAS were able to process contained an unusual number of errors. For example, a student from Beloit College in Beloit, Wisconsin was listed as attending a college he had never attended, failing courses he never took, and having been disciplined by university officials in 1967, even though he was only 9 years old at the time.

Michael Rappaport, assistant dean of admissions at UCLA Law School told The Chronicle of Higher Education, that he's "noticed a lot of mistakes-more than usual. We are beginning to wonder if we are missing things."

Sandra G. Mannix, director of admissions at Villanova University's law school also reported finding numerous errors. For example, one applicant's score should have been 100 points lower than was recorded by ETS. Before the mistake was corrected, the applicant had been accepted by Villanova and several other law schools.

Admissions officers at the Temple University law school ran a spot check for errors in January on 80 reports. They found errors in 20 per cent of them.

Students currently pay $15 to take the Law School Admissions Test and an additional $38 to have their transcript and computerized summary sent to five law schools.

(Testing Digest, Spring 1980, p. 15)

Appendix C—Tamar Pachter Article, $450,000 Fund Article—Testing Digest, Spring 1981

ETS Tests Researchers' Wiles

Many people believed that when New York state passed its Truth in Testing bill in 1978, it had lifted the cloak of secrecy surrounding the **Educational Testing Service** (ETS) and other testing institutions. The New York law, which requires standardized testers to provide full grading and test question disclosure, did open up the floodgates of legislation. About twenty states have now passed some form of open-testing legislation, and some of these bills are more strict than the New York prototype. But it appears that all this lawmaking, rather than opening the industry to closer scrutiny, has served to make ETS in Princeton an increasingly guarded bastion of testing. ETS is no more willing now than it was last spring (see "ETS and the 'Coaching' Cover-Up," *NJM*, March '79) to share the fruit of its years of catalogued studies on the effects of coaching on aptitude test scores, research which some independent investigators have called into question.

Just recently, two outspoken critics of standardized testing say they were denied admission to the ETS library archive, despite their having made appointments to study its collection. The archive is rumored to contain not only early ETS history, but also the results of studies ETS cites as evidence that coaching does not improve test scores. Evidently, ETS did not consider it to be in its interest to allow **Steve Solomon,** testing project director for the New York Public Interest Research Group, and **John Weiss,** codirector of the Committee for Fair and Open Testing, to examine the archives. "The archive is only open to responsible researchers," Weiss says he was told by a librarian. The official ETS position has a similar ring: "Those two edit a paper which is the *National Enquirer* of testing," says **Mary Churchill,** the public information officer at ETS. The paper, actually a newsletter of the Committee for Fair and Open Testing, "contains slanderous things about ETS," adds Churchill. "As far as those two are concerned, the archive is not a matter of public information."

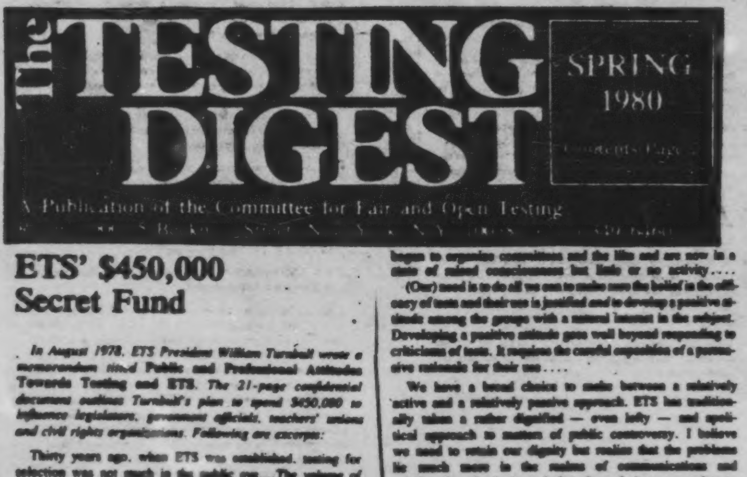
But Solomon and Weiss are not the only researchers who recently have had trouble getting their hands on information in the ETS archive. When Drs. **Warner Slack** and **Douglas Porter** of Harvard Medical School tried to review results of ETS research on coaching—including the Pike Study, which indicates coaching can work—for a study on the SAT published in the May edition of the *Harvard Educational Review*, they also were thwarted. "People were very cagey about the Pike Study," says Dr. Porter. "They said that they would send it to us when it was ready, but of course, it was already ready." Porter says that he finally received the controversial study, prepared in 1972, from a member of the College Board, an ETS client.

Slack and Porter's investigation—based on ETS's own published research—concludes that the SAT should be eliminated as a requirement for college applicants because it is highly susceptible to coaching and is not a test of aptitude. Wrote Slack and Porter: "In publications sponsored by ETS or the College Board, we found nine conclusions that studies to date had demonstrated coaching to be ineffective, even though the published experimental evidence indicated the contrary. We also found seven assertions that gains in scores from coaching were not significant, although they were statistically significant."

While ETS freely admits that the coaching issue merits further study, it says that this research will be conducted in-house. One wonders, though, how seriously college admissions officers should take the studies of an institution that is so unwilling to provide its critics with uncensored information.

—*Tamar Pachter*

A copy of the entire 21 page confidential memo was attached to my testimony The following is a portion of the memo that was reprinted in the Spring 1980 Testing Digest



The TESTING DIGEST

SPRING 1980

Contents Page 2

A Publication of the Committee for Fair and Open Testing

[illegible]

ETS' $450,000 Secret Fund

In August 1978, ETS President William Turnbull wrote a memorandum titled **Public and Professional Attitudes Towards Testing and ETS**. *The 21-page confidential document outlines Turnbull's plan to spend $450,000 to influence legislators, government officials, teachers' unions and civil rights organizations. Following are excerpts:*

Thirty years ago, when ETS was established, testing for selection was not much in the public eye....The volume of criticism was turned up in the '50s and '60s by Banesh Hoffmann's writing on the *Tyranny of Testing* theme castigating objective tests for their alleged failure to reward deep thinking....In the 1960's, with the 20-fold expansion in use of tests for selection...(w)e grew to be a large conspicuous organization....

Accordingly, by the early '70s ETS stood as a powerful, little-known organization that many people had to deal with involuntarily in order to gain (or be denied) access to educational opportunities. We provided scores that were mysterious not only to students and parents but also to the educators to whom they might turn for guidance.

This situation ushered in the era of angry, hostile, and sometimes paranoid attacks on testing and on ETS. The NEA 'moratorium' resolution came out in 1972. At about the same time, the Association of Black Psychologists and other minority groups articulated their anti-testing positions....The first clear expression of the mood of an extremist fringe against ETS *per se* was probably the Brill article of 1974—only four years ago.....Since then, the shrill tone has been carried forward in the PIRG attacks, which quote Brill extensively, in Nader's appearances (e.g., APA 1976), in the NAESP journals and books.

Associations like APA, AERA, and NCME, and other groups professionally concerned with testing and its use, had been passive observers of developments through the early '70s, but by about 1975 they began to see that the hostility had wider ramifications and deeper significance than they had thought when Hoffmann was tweaking testing's tail. They began to organize committees and the like and are now in a state of raised consciousness but little or no activity....

(Our) need is to do all we can to make sure the belief in the efficacy of tests and their use is justified and to develop a positive attitude among the groups with a natural interest in the subject. Developing a positive attitude goes well beyond responding to criticisms of tests. It requires the careful exposition of a persuasive rationale for their use....

We have a broad choice to make between a relatively active and a relatively passive approach. ETS has traditionally taken a rather dignified — even lofty — and apolitical approach to matters of public controversy. I believe we need to retain our dignity but realize that the problems lie much more in the realm of communications and politics than in logic and that the solutions are going to be found (or the battles lost) in those hot arenas rather than in cool reason.

In our communications to non-psychometric groups, we should cite fewer statistics and more human interest cases.

Within the context of an active approach combining both logical and political components, we should:

1. Attempt to improve the situation and let people know what we are doing, and
2. Increase and improve our communication with non-test publics: friendly, hostile, and in between.

Each of these two lines should be pursued partly by ETS on its own and partly in cooperation with one or more other agencies. Accordingly, we would have a fourfold approach: improvement and communication, alone and in partnership.

Within each of the key groups identified as critical of testing (minorities, students, education groups, social critics) there are elements that in fact are well disposed. We should seek out and develop alliances with those elements....

Potential Projects

In discussing strategy, I suggested two lines of action:

1. Attempts to improve the situation and
2. Efforts to communicate to the general public.

....In Display F (see) the kinds of projects that might develop.

(continued on next page)

Display F
Possible Activities with Public Impact
(10 of 19 Projects listed)

- Explore the possibility of establishing a 'test use review board' (under a better name) to which complaints of improper, unprofessional, unethical practice can be referred.
- Develop a program of regional and state workshops after the style of joint workshops with NCME. Seek financing from NIE. Look toward holding at least 30 in the next two years. Could be entirely with NCME or might broaden to work with AERA, AMEA, possibly APA.
- Commission a good analytic poll by Gallup to discover what various subgroups of interest think about tests and testing.
- Work with Chief State School Officers (and State Testing Leaders?) on development of alternative models for design of statewide assessment—especially basic competency—and for reporting of scores to schools and parents (PTA might be involved).
- With AFT, develop recommended models for teacher evaluation, including proper role of tests of teachers and pupil gain data.
- With Consortium, develop a paper on Alternative Evaluation Systems and a training program for school people on how to use them.
- With ACE, convene a group to examine implications of governmental encroachment on institutional and professional authority in decision-making about admission and other student-related topics. Ranges from Bakke to Harrington, and includes privacy (Buckley), some provisions for Handicapped, graduation criteria. Could be cosponsored by client groups.
- Convene a conference on self-policing by the profession, looking toward definition of areas of concern, present mechanisms, gaps, needed action.
- Work with legislators' offices on the shaping of bills affecting testing or policy based on results of assessment. Do some of this work *pro bono*.
- D.C. Office expansion. This is under way. We should be sure to give it the fullest support.

It may be that we are witnessing a long-term need to step up our communication efforts. In any case I'm inclined to the view that we ought to see the next 3 or 4 years as a time of unusual need and opportunity. Some of the actions to be taken can be supported under existing budgets for Information Services, Advisory and Instructional, or R&D, but some of them probably cannot and should not be so supported. Accordingly, I think we should anticipate an extraordinary allocation of funds out of ETS net for this work in 1978-79 through 1981, on the order of $200M, $150M and $100M for the three respective years. This would play hob with Mr Solomon's formulas for allocation of net.

William W. Turnbull
August 14, 1978

Abbreviations Used in Turnbull's Memo

ACE - American Council on Education

AERA - American Education Research Association

AFT - American Federation of Teachers

APA - American Psychological Association

Brill Article - "The Secrecy Behind the College Boards" by Steven Brill, *New York Magazine*, October 4, 1974.

Consortium - National Consortium on Testing

Harrington Bill - Congressman Michael Harrington (D-Mass) introduced a Truth-in Testing bill in 1977.

NAESP - National Association of Elementary School Principals (See the July/August 1975 issue of *Principal Magazine*).

NCME - National Council on Measurement and Evaluation

NEA - National Education Association

NIE - National Institute of Education

NSBA - National School Boards Association

PIRG - Public Interest Research Groups. Student funded research and advocacy organizations.

Mr. Solomon - Robert Solomon, ETS Executive Vice President.

Tyranny of Testing - Book published in 1962 by Dr. Banesh Hoffman; noted mathematician and physicist.

The TESTING DIGEST
Spring 1981
A Publication of the Committee for Fair and Open Testing

Question: "Mrs. Wilson is ready to go home. She has worked all day and will work at home tonight. She is taking some books and papers home with her. Mrs. Wilson is probably:

a) a waitress in a restaurant,
b) a schoolteacher
c) a store clerk.

—*Stanford Achievement Test, 1973*

Note that the choices do not include lawyer, executive, sales manager, or professor. Even the fact that she works outside the home makes this an unusual test item.

Sex Bias in Testing

by Phyllis Rosser

Two important mid-seventies studies were inspired by the passage of Title IX of the Education Amendments (which said it was illegal to discriminate against women in testing or counseling). One of them looked at eight of the most frequently used achievement test series and found women and girls portrayed in stereotyped "home-making" roles, passively watching boys play or work while men were actively involved in a variety of jobs or professions. Male nouns and pronouns outnumbered females in all but one test series, sometimes at a ratio of 14 to one. The second study looked at 5th, 8th and 10th grade tests to see if sex-biased content had an effect on girls' scores and found that girls tend to do better on questions with more females, equal numbers of males and females, or people of unspecified sex in them than on questions that featured males. Since females were less frequently mentioned in these tests, the implication was clear—not only were they presenting a demeaning and unrealistic picture of females in society, they were also unfair measures of achievement. With these studies in mind, I examined several widely used standardized achievement, aptitude and intelligence tests.

Phyllis Rosser, a contributing editor at Ms. Magazine, writes on educational issues.

Achievement Tests

In late spring most children take some form of achievement test to calibrate how much they've learned over th course of the school year and determin their level or track for the following yea Achievement test scores are also used t check teachers' effectiveness and to compare a school or school district with it others.

The following widely used tests published between 1971 and 1974 were reviewed Iowa Test of Basic Skills published b *Houghton-Mifflin;* California Achievemen Test published by *CTB/McGraw-Hill;* and The Metropolitan Achievement Test an The Stanford Achievement Test, both o which are published by *The Psychologica Corporation, a subsidiary of Harcourt Brac & Jovanovich.*

Each of these tests portrayed wome in subservient and/or demeaning situations. Women were shown in the traditional roles of homemaker, teacher, nurse and babysitter, and girls helped mothe cook, wrote to pen pals or watched passively while men and boys led active interesting and varied lives. Most gende specific items on these tests used mal examples. For instance, on the Iowa Tes of Basic Skills, male nouns and pronouns outnumbered female nouns by a ratio o nearly 5 to 2.

Revisions of each of these tests published between 1977 and 1979 show significant egalitarian gains. Males an

(please turn to page 3

Inside...

"Culture-Fair" I.Q. Tests 2
Subscription Blank 3
Testing in China 6
Adult Competency Testing 8
Chuck Stone on ETS 10
LSAT items 13
Students Outwit ETS 14,16
Boston Conference—
May 2nd, 1981 15
Truth-in-Testing Testimony 16

The 'Culture-Fair' WISC-R I.Q. Test

In 1969 the American Black Psychological Association called for a ten-year moratorium on the I.Q. testing of minority children. By 1974 the NAACP and the National Education Association were both publicly urging schools to stop administering tests to children of all colors. Reacting to these pressures the Psychological Corporation, then an independent company and now a subsidiary of Harcourt Brace Jovanovich, revised their Wechsler Intelligence Scale for Children (WISC) to create the culture-fair Wechsler Intelligence Scale for Children - Revised (WISC-R). The WISC-R was standardized on a population which was 50% male, 50% female, 85% white and 15% Black, Latino, Asian or Native American. Today the WISC-R is America's most widely administered I.Q. test.

Last October Northern Californian Federal Judge Robert Peckham ruled that the WISC-R as well as the other commercially available I.Q. tests were biased, and could therefore no longer be used by the public schools in his district jurisdiction to track minority children.

In December 1980 an Illinois federal judge ruled that the WISC-R test was not biased and could be used to track children. In an effort to bring some empirical evidence to this dispute, let's examine several questions from the culture-fair WISC-R IQ test.

1) What is the thing to do when you cut your finger?
2 point response: Put a bandaid on it
1 point response: Go to the hospital
0 point response: Cry, bleed or suck on it
—WISC-R IQ Test

Minority children usually perform poorly on this item. A few years ago a Baltimore, Maryland sociologist asked several inner-city youths why they answered the question the way they did. She found that many of these kids answered "go to a hospital" because they thought that cut meant a big cut. Or they thought it was a small cut—and since they didn't have any bandaids at home, they answered "sucked on it"—and received no points.

2. What is the thing to do if you lose a ball that belongs to one of your friends?
2 point response: Buy him a new one and pay for it
1 point response: Look all over for it, try to find it
0 point response: I'd just cry, tell him you're sorry, apologize
—WISC-R IQ Test

If you're poor, it often is not possible to buy another one and it is unlikely that you have one to give. If the ball is really lost, looking for it is a waste of time. And if it is not really lost but could be found, wouldn't looking for it be more "intelligent" than buying a new one? Crying or apologizing might be the best response for a poor child, but such an answer will yield a lower "intelligence" score.

3. What are you supposed to do if you find someone's wallet or pocketbook in a store?
2 point response: Return it
1 point response: Give it to the store owner
0 point response: Keep it; make believe you didn't see it
—WISC-R IQ Test

Is a poor hungry child really less intelligent if s/he keeps the money? Also a child—especially a minority child—might ignore the wallet/pocketbook for fear that s/he would be accused of stealing it. Such a response would earn the child no points.

Other WISC-R items include:

- What are some reasons why we need policemen?
- Why is it usually better to give money to a well-known charity than to a beggar?
- From what country did America become independent in 1776?
- Who discovered America?
- What is the capital of Greece?
- What are hieroglyphics?
- Who was Charles Darwin?

—J.W.

——ETS on Test Bias——

In 1970 ETS reacted to charges that their tests were culture-biased by making one out of every eight reading quetions minority oriented. *New York Magazine* reporter Steve Brill's interview (10/7/74) with ETS senior psychometrician Marion Epstein about the addition of the minority-oriented section follows:

Brill (Q): "If the test weren't culturally biased in the first place, why did you make the change?"

Epstein (A): "Because minorities feel at ease reading this kind of passage."

Q: "If they feel at ease reading this one, does that mean they don't feel at ease reading the six or seven other passages in the test?"

A: "No. It just means they feel more comfortable with this one."

Q: "Well, if they feel more comfortable, does that mean their scores will be higher?"

A: "No. I don't think there will be any difference in scores."

Q: "Well, if there won't be any difference, why would you make the change. Was it just so you could look like you were doing something?"

A: "No, it's because when people are more comfortable, they'll do better on the test. They feel less threatened."

Sex Bias

Continued from page 1

females are mentioned in equal numbers. Girls are shown participating in such previously boys-only activities as bicycle riding, fishing, cutting the grass, delivering newspapers, handling money and even winning science awards. Women were shown driving, not only cars, but even cabs and trucks, shooting rapids in a canoe and travelling in space.

> *"Women are never depicted as inventors, doctors, lawyers, politicians or in a position of power higher than a man's within a question."*

Women for the first time are allowed to enter the professions as dentists, veterinarians, even scientists—including a black woman on the California Achievement Test. Women, however, are never depicted as inventors, doctors, lawyers, politicians or in a position of power higher than a man's within a question.

But these new improved tests are not widely used. Most school districts save money by buying answer sheets for the older, stereotyped test books they already own.

I.Q. Tests

Measuring the intelligence quotient or I.Q. of an individual is a mysterious activity because no one knows exactly what the components of native intelligence are. Nevertheless I.Q. tests are used to label children as "smart" or "stupid" at a very early age—affecting teacher expectations, placement in class, and ultimately, determining a child's future. Two of America's most widely used I.Q. tests were reviewed: the Wechsler Intelligence Scale for Children–Revised (WISC-R), administered to children between ages of 5 and 16, and its companion test for persons 17 years of age and older—the Wechsler Adult Intelligence Scale (WAIS). Both of these tests are published by the Psychological Corporation, a subsidiary of Harcourt Brace & Jovanovich. These exams are written with the egalitarian assumption that men and women's average intelligences are equal. Harcourt Brace accomplishes this by statistically balancing the test so that each sex will answer approximately the same number of questions, correctly though not the same ones.

The message one gets from looking at the WISC-R and the WAIS is very clear—it's a man's world. In a general information section, children are asked questions about Columbus, Edison, Darwin, while adults are overwhelmed by references to such famous or high status men as Longfellow, Hamlet, Goethe, U.S. presidents ("name four since 1900".) Yet not a single famous woman is mentioned in this section.

The math sections, where girls traditionally perform poorly, are even worse. On the WISC-R there is one first grade level question about a girl losing some ribbons while many of the other questions show boys and men buying, selling or working. On the WAIS math section, there are no females mentioned in any of the fourteen items.

On the WISC-R picture arrangement section (putting a series of pictures in order to tell a story—which purports to measure a person's ability to analyze complex situations) 10 of the stories feature males, only three involve females. In a sequence titled "The Flirt" a king asks his driver to stop the car so he can flirt with a beautiful woman. The correct answer has the sequence end with them strolling down the street together. Another sequence features a man with a female department story "dummy."

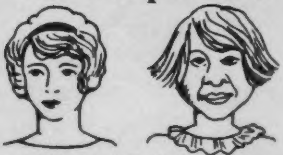
The content of these tests raises several important questions: Does the WISC-R

(continued on page 4)

———Attention———

We hope you find this complimentary copy of* The Testing Digest *informative. Subscribe Today!

Which is prettier?



FROM THE STANFORD-BINET IQ TEST
USED TODAY TO "TRACK" ELEMENTARY SCHOOL CHILDREN

THE TESTING DIGEST

"Indispensible for anyone who wants to keep up with the important activities in testing."

Harvard Medical School Professor Douglas Porter,
Co-author, *"The Scholastic Aptitude Test: A Critical Appraisal"*

☐ Time Magazine called our last issue "provocative". ETS called it "The National Enquirer of Testing". We provide timely information to those legislators, parents, teachers, students, and scientists who are active in the minimum-competency, IQ, "Truth-in-Testing", and employment/licensing test reform movements.

Subscribe!

The Testing Digest
Published quarterly. Check one.

☐ Students $4.50 ☐ Institutions $18.00
☐ Employed $9.00 ☐ Single Copy $3.00
☐ 1979 and 1980 issues: $18.00

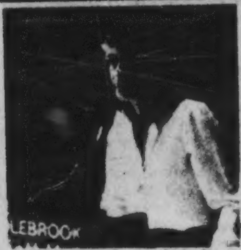
Name __________

Address __________

City ______ State ___ Zip ____

Mail to: *The Testing Digest*, Room 1000-C
5 Beekman Street
New York, N.Y. 10038

Enclosed is $ ______



"Women do well on standardized tests to the extent that they identify with the male culture."
—Phyllis Rosser

Sex Bias . . .

Continued from page 3

and WAIS' male content have an effect on female tests scores? Does the demeaning content have an effect on female self-confidence during the test—and in later life? And why must an I.Q. test contain material that is offensive to women?

Aptitude Tests

Aptitude Tests—which attempt to measure a person's potential—are used by vocational counselors and by college and graduate school admissions departments. One of the most popular counseling exams—the Differential Aptitude Test published by the Psychological Corporation, a subsidiary of Harcourt Brace Jovanovich—contains a section on mechanical reasoning where boys score much higher than girls. The content is highly male-oriented: there were pictures of men looking in mirrors, riding horses, throwing a ball, using a hose, and standing on the moon. The only woman on the test is sitting in a wheelchair. No famous women appear in the Verbal Reasoning section although there are two questions featuring famous men. In the Language Usage section, male nouns and pronouns outnumber females three to one. Men and boys are involved in important activities while girls are careless, break their dolls or are part of trivial comparisons.

Sex balancing becomes critical on aptitude tests when college entrance is at stake. On the Scholastic Aptitude Test (SAT) which is administered by the Educational Testing Service (ETS) and used by most selective colleges for admissions purposes, girls and boys have the same average verbal scores. Girls used to score higher on the verbal section, but as ETS researcher Carol Dwyer explained "the verbal section has been rewritten so that boys and girls now score equally well."

While the verbal section has been "balanced" for sex, no similar equalizing of scores has occurred on the SAT's math section where boys outscore girls by an average of 50 points. In fact in the past few years, ETS has eliminated several algebra and data sufficiency questions (determining whether there's enough information to solve a problem) which usually favored girls. ETS researcher Thomas Donlon observed in one study of the SAT's match section that his company's psychometricians had the power to vary the scores between the sexes by as much as 60 points depending upon the questions they selected. Dwyer says the decision to "equalize" the verbal section happened gradually over a period of 20 years. According to Donlon it reflects a shift in the interests of colleges from liberal arts to more emphasis on technical subjects and engineering. This shift in interests undoubtedly affected content. As Donlon noted, in a 1975 study on the SAT, "If the thirty science and practical affairs items had been matched in number to equal the twenty-four aesthetic, philosophical and human relations items, women's SAT verbal scores might have been higher than men's scores."

Questions in the ETS guidebook, *Taking the SAT*, seem to emphasize power and accomplishment. English questions refer to "rulers and administrators who hold the world in their hands," "man's aspirations and wonderments," and individuals who believe that they have the right to trample over ordinary people. Famous men (Shakespeare, Dylan, Hitler, Freud, Napoleon) far outnumber famous women, even in questions about the arts—a content area that's supposed to favor women.

Male nouns and pronouns predominate in all sections of the sample test (no females are mentioned in the math sections although men's names are used in three questions). Male occupations outnumber females (who are mainly writers) almost 3 to 1, but there is a woman mayor in the English grammar section.

A report published by a College Entrance Examination Board commission openly admits that "the SAT scores tend to reduce the advantage girls enjoy in graded school work." ETS researcher Marlaine Lockheed raises a critical question—"if the test contains more female-oriented content, will male test scores go down?" Since girls' college grades tend to be better than boys, we know the SAT doesn't predict how well they will do in college even though this is the purpose of the test.

Most colleges and professional schools use a mathematical formula to combine grades and aptitude scores into a single admissions number that is supposed to predict first year grades. According to a 1973 study by the American College Testing Program (an ETS competitor) the equations most universities use tend to systematically under-predict girls' actual performance in college. Nancy Cole, Director of Educational Research at the University of Pittsburgh, believes that if colleges used a formula that fairly predicted girls first year grades, "they'd be accepting two-thirds girls to one-third boys."

On the graduate level, the Law School Admission Test (LSAT) sample book features questions testing "data evaluation," "data application" and "logical reasoning ability" that often have a business and economics content. Men predominate throughout the test and are shown as presidents, senators, famous writers and political thinkers. Not a single famous woman is mentioned and no women are shown in a position of power or even working in a profession. Only one woman, a partner in a family business, is allowed to make business decisions. And another

(please turn to page 3)

The Testing Digest **Vol. III, #2**
Co-Directors: John Gordon Weiss Steven Solomon
Associates: Debra Ettinger Lawrence Kronick
Published by The Committee For Fair & Open Testing
Room 1000, 5 Beekman Street, New York, NY 10038

Continued from page 4

woman, referred to as "a businesswoman with a reputation for honesty," is involved in a slander suit against her neighbor.

Several questions seem geared towards disrupting a woman's self-confidence. For instance, this item reminds law school applicants that women have been regarded as property for a long time: "In Roman law, a woman was known first as her father's daughter and later as her husband's wife: she had no identity of her own."

Or this LSAT reasoning ability item where test takers must choose/select the answer that most nearly resembles the question: "Do you think that our university ought to go on discriminating against disavantaged students by continuing its current admission policies?" The answer—"Should your neighbor stop beating his wife?" This seems like a double whammy for minority women.

Attornies Teri Roth and David White report in Harvard University's *Women's Law Journal* that the women who take the LSAT are superachievers. They have extremely high LSAT verbal scores, higher college grades and will have higher law school grades than the average male LSAT test taker. Yet male and female LSAT score averages are identical. If the purpose of this test, like the SAT, is to predict success in school, then it doesn't work as well for women as it does for men.

"The only woman on the test is sitting in a wheelchair."

Over the past decade White reports that the LSAT "has increasingly emphasized mathematical ability even though math is not a high priority skill for lawyers." White continues "that since women usually have less interest in math than men, the LSAT's math orientation puts women at a disadvantage." Perhaps as a result of White, Roth and others' critiques, the Law School Admissions Council recently announced that starting in mid-1982 they will eliminate some math questions from the LSAT.

□ □ □ □

Eliminating sex bias in tests that affects both girls' test scores and their self-confidence seems like a matter of simple justice—providing girls with equal opportunities for success in life.

But this doesn't deal with the power which test makers have over our lives in a larger sense.

Depending upon the test publishers' hidden or not so hidden agendas, different groups of people will be labelled as intelligent, achievers, or having aptitude. Currently people with math and science talent are likely to score highest on tests. The answer, however, is not simply to give more weight to verbal questions. Instead Americans must realize that since standardized tests can be so easily manipulated, it is impossible to make them completely fair. We need to accept this fact and devalue test scores instead of allowing them to be the major determining factor for getting ahead in life.

—Types of Bias—

Studies done on I.Q., achievement and aptitude tests have identified five types of potential sex bias.

- The most blatant type is the demeaning and stereotyped depiction of women and girls on many tests.
- Another common type of bias is that males are more often featured in test questions. A 1975 Educational Testing Service (ETS) study found this has a direct effect on achievement test scores. Girls are more likely to get a question right if there are more females, an equal number of males and females, or people of unspecified sex within the question.
- A potential source of test bias results from difference in interests and skills which emerge in junior high school and follow the lines of traditional socialization. Several studies have shown that girls perform better on math questions that involve "typical" female activities. For example, girls answered a question about feet and inches correctly more frequently if it was put in the context of drapery making than the heights of male basketball players. This source of bias deserves more than the small amount of research that now exists because poor scores may now be discouraging girls from entering these traditionally male-oriented occupations.
- Girls score higher on questions involving the arts and humanities while math, science and business oriented questions favor boys. Girls excel in reading comprehension, algebra and biology, while boys' vocabulary, analogies, geometry and physics scores are higher.
- The sex of the examiner may also contribute to bias. One study indicates that girls score higher on the WISC-R arithmetic section when tested by a female examiner and boys score higher when tested by a male. Since most clinical and school psychologists are male, this might be lowering girls scores. Harvard psychologist Robert Rosenthal reports that "the sex and interaction of the testers have significant effect on results. Even though they're not supposed to give any feedback, "a warmer examiner will get a performance that is superior to a colder one," he says.
- The rushed nature and multiple-choice format of most standardized tests may be other sources of bias. Initial research indicates that on multiple choice tests, women answer a higher percentage correct than men, but have fewer completions. Therefore it is probable that if test takers were given more time, women's scores would go up. There is also some evidence that essay questions improve girls' scores. Over the past two decades, however, essay questions have been deleted from tests because they can't be scored by a computer.

Women at ETS

Marlaine Lockhead, the newly appointed Chairperson of the Committee on the Status of Women (CSW) at ETS, announced the completion of a recent study entitled, *The Status of Women at ETS—Compensation and Clarification.* "One of our major findings is that the average salary of women at ETS as of July 1, 1980, was 60 percent of the average salary of men at ETS. That's close to the national figure of 59 percent and reflects the higher concentration of women at lower salary grades," stated Lockhead. ETS has deferred any action on this study pending the completion of another study.

Special Report

China

Cheating in China

By Bonnie Sexton
Peking, China

An entire high-school graduating class has been charged with acting in collusion with a teacher and an assistant principal to cheat on university entrance examinations, according to a report in *China Youth News*, a national newspaper for young people published here.

The case has dramatized the intense pressures on students to get high grades, in a country where very few qualify for admission to universities. This year, for instance, only 300,000 of the 5 million students who took the examinations were admitted.

The cheating incident took place in Henan Province. According to the newspaper account, the assistant principal and the teacher told the students to disregard rules on the examination, to help each other with the answers, and to bring reference materials to the classroom. When students said they were afraid they would be caught, they were told that "the examination proctors' reports of cheating will never leave the province."

As a consequence, the students brought notes and books to the test rooms, worked with each other to formulate answers, and took a longer time than was allowed.

Eventually, word of the cheating did reach higher authorities, apparently from a proctor who said he had been threatened by the assistant principal and some of the students.

As a result, the examination results were declared invalid and the students were told they would not have a second chance to take the tests. The assistant principal was fired and the teacher arrested.

In an editorial that appeared along with the account of the incident, the newspaper blamed the cheating on the fact that the Cultural Revolution had weakened respect for law and order and had concentrated power in the hands of a few people.

The treatment of those involved in the incident is harsh by Chinese standards, since dismissal from a job is a severe and rare punishment and the students may have been permanently denied a college education.

There are other reports of pressure on students to do well on the entrance examinations, and on teachers and school administrators to prepare them for tests.

Parents in Peking are reportedly asking for longer school hours in the weeks before the examinations are given, and one foreign teacher in southeast China said that physical education, art, and music had all been dropped for several weeks before the examinations so students could devote more time to preparation.

Bonnie Sexton writes for The Chronicle of Higher Education.
© The Chronicle *1980*

PRC Professors Visit ETS

ETS recently hosted two professors from the University of Peking Foreign Language Institute's English Department. The Chinese team's primary goal was to learn how to score and construct the Test of English as a Foreign Language (TOEFL), a test that most U.S colleges require foreign applicants to pass.

Professor Ding Wardoo told *The ETS Examiner* that "In the People's Republic of China our methods are somewhat conservative and we want to improve." He continued that today standardized multiple-choice tests are rarely, if ever, used in the PRC. If details can be finalized, a PRC research team will travel to ETS later this year to study "the whole [testing] process and also to learn about ETS' International Instructional Programs," Ding concluded.

The Committee has written to ETS, the Chinese Embassy in Washington, and the University of Peking English Department requesting a meeting with the PRC's testing team. If you would like to attend this meeting please contact The Committee.

—And You Think You've Got It Tough—

"The examinations usually took place at permanent quarters which were created in different cities especially for this purpose. The halls were arranged in alleys and further disposed in rows of cells—each alley being named and each cell numbered. The cell was about four feet by three, with the height of a man, and open on but one side. Grooves were made in the wall to admit a plank, which served as an eating table or a writing desk by day and a bed by night.

"Before entering the examination hall, all the candidates had to be strictly inspected: their robes, pockets, shoes, inkstones, and other personal belongings being searched, lest "skinning paper" or "cribbing compositions" be smuggled in. Each examination consisted of three sessions, lasting for a period of about three days. In the name of the emperor, the chief examiner called upon the spirits to inspire the minds of the candidates according to their ancestral merits. As he intoned this prayer many firecrackers were set off. Thus the examination began.

"During the three days of confinement in the cramped cells neither examiners nor candidates were allowed outside communication. No book or written paper was permitted to be carried upon any person. Names were sealed and each paper was marked with a cipher; thus no examiner could discover the possible identity of the candidates. Strict regulations concerning the examination and severe punishments for offenders were enacted so as to guarantee honor, spirit, and fair play. A single instance of cheating in the examination might lead to wholesale execution of all involved, examiners and candidates alike."

From a description of civil service examinations in China in the 14th Century, The Princeton University Library Chronicle, Summer 1959, page 178.

A fully-documented exposé of the standardized testing currently used for college, the professions, civil service, business, and industry

The Testing Trap

ACCLAIMED WITH ENTHUSIASM

"An important book which judiciously tells us what is wrong with standardized testing and the testing industry... It is a book which should be widely read."
—ASHLEY MONTAGU

"A level-headed, informative assessment of testing that will be useful to anyone concerned about education, and invaluable for parents and students who face encounters with standardized tests."
—JAMES FALLOWS, Washington Editor, *The Atlantic Monthly*

"An invaluable service to the many people victimized by tests... It should motivate Americans to open up the testing industry."
—ALTHEA T.L. SIMMONS
Director, NAACP, Wash., D.C.

"Strenio has captured the essence of one of the most complex educational issues of the 80s...and has set a course of action for much needed change."
—REBECCA FISHER,
National PTA

THE TESTING TRAP, 360 pages, fully indexed. At bookstores $14.95, or send check (plus $1 for postage and handling) to Rawson, Wade Publishers, Inc., 630 Third Avenue, New York, NY [illegible]

RAWSON RW WADE

> "One out of five American adults lacks the skills and knowledges (sic) to function effectively in the basic day-to-day struggle to make a living and maintain a home and family."
>
> *HEW News, 10/29/75*

by Lawrence Kronick

What led the Department of HEW to issue such a proclamation? If true, surely such a state of affairs is as worrisome as an unemployment rate of 20 percent. For this news release says not just that adults lack jobs or lack skills for jobs, but that they lack the prerequisites to function effectively in life. How did HEW reach this ominous conclusion? This question is explored in a recent NIE report entitled "The APL Study: Science, Dissemination, and the Nature of Adult Education," by Huron Institute researcher Walt Haney and Lloyd David of the Continuing Education Institute. A synopsis of their findings follows.

Reacting to a dramatic increase in adult educational programs the Adult Division of the U.S. Office of Education (USOE), in August 1970, issued a Request for Proposals (RFP) for a research project that would establish "the functional competence needed for meeting the requirements for adult living." USOE officials hoped that once "the requirements for adult living" had been defined, educators could use these specifications to teach adults the skills they needed to function in modern America.

In the summer of 1971 USOE awarded the Texas State Educational Association (TSEA) a million dollars to complete this project. The TSEA then subcontracted with the Division of Extension of the University of Texas at Austin to actually develop "the requirements for adult living." The University of Texas researchers on "the basis of 'a review of related literature and research,' decided that a person's success in life correlated positively with his/her years of schooling, income level and occupational status." The Texas researchers then developed a list of basic skills and general knowledge areas that successful people possessed. From this list test items were developed and then administered to 4,000 adult volunteers, most of whom were participants in adult education programs in 30 states.

The Texas State Education Association reported to the USOE that "the most important product" of the University of Texas work was a listing of the "requirements for adult living which we have tentatively identified as contributing to success in adult life."" The report concluded: "For the first time, there is a body of research which offers guidance on what we should be teaching our students in Adult Basic Education."

In 1973, an Adult Performance Level Test consisting of a set of adult competency test items based on the Texas research, was administered to a representative sample of 7,500 adults throughout the continental United States. The HEW's Final Report on Adult Competency Education Profile, based on the results of this test concluded: "In general, the answers to the questions posed by this section is not as competent as we thought."

Overall, one out of five (19.8%) adults were said to lack skills and knowledge needed to function effectively, another 33.9% were rated as marginally competent, and only 46.8% were said to be functioning with any degree of real competence.

The report concluded that more than 40% of Black and Spanish surnamed adults were "functionally incompetent."

This study and its findings roused considerable attention. The title of Edith Roth's lead article in the May 1976 USOE-published journal *American Education* credited the study with causing "A Ferment in Education." She wrote:

> *Today—not quite five years since the APL study began—more than 30 states are either gearing up to teach or are already teaching adults to gain competency in the 65 practical objectives which the University of Texas team developed.*

In March, 1975 the APL approach also received the seal of approval from the Joint Dissemination Review Panel (JDRP) in HEW's Education Division. Just a few days after the JDRP's approval, USOE and the University of Texas at Austin entered into an agreement with the American College Testing (ACT) Program under which ACT received the "exclusive right to refine, adapt, publish, and distribute the APL materials."

APL Revisited: Its Uses and Adaptation in States *can be obtained from Jerome Lord, Adult Learning Team, National Institute of Education, 1200 19th St. N.W., Washington, D.C. 20208, (202) 254-5706.*

By the end of 1976 there were at least 120 APL based adult competency educational projects in 30 states. In addition several states used APL items in their minimum-competency tests that high school students had to pass to receive a high school diploma.

In 1976, the USOE cited the APL project as an exemplary educational program, and Congress authorized 10 million dollars to subsidize states' Adult Basic Education Programs. The USOE then gave priority for funding to those

(continued on page 9)

—APL Item—

Mary Dilly is a housewife. She needs to make some extra money for a while. She can type, bake, and sew. The quickest thing for her to do to make money is to:

a. Apply for temporary office work.
b. Go to school.
c. Bake a cake.
d. Make her own clothes.
e. I don't know.

— *The Adult Performance Level Study, 1973*

—APL Item—

The city garbage truck has not picked up Esther Maxey's garbage of three weeks. Esther is having trouble keeping the flies and mice away. What should she do?

a. Take the garbage down the street to an empty lot.
b. Call the hospital to complain about the mice.
c. Call the sanitation department about the problem.
d. Cover the garbage with a sheet.

American College Testing Program, 1978

Continued from page 8

states that implemented programs based on the APL research. In addition, HEW spent additional monies sponsoring numerous workshops on the APL program. All these factors combined resulted in the widespread utilization of the University of Texas researchers' work.

In 1976 the University of Texas researchers received a grant from the State of Texas to build a curriculum to complement the APL test. Three years later the University of Texas at Austin entered into an agreement with Harcourt Brace & Jovanovich to market the curriculum program entitled, "The APL Series." Today ACT's APL Test and Harcourt's APL Series are used in hundreds of Adult Basic Education Programs nationwide. In addition APL test items are used by at least a dozen states in constructing the minimum-competency tests high school students must pass to receive a diploma.

* * *

The preceding account reveals how an $11 million plus government subsidy combined with entrepreneurial initiative can result in the implementation of a testing program in nearly three dozen states and the widespread dissemination of a conclusion that nearly 20 percent of Americans were functionally incompetent. Haney and David spend the bulk of their report analyzing the forces that promoted the APL program and summarizing other researchers' criticisms of the APL test. Following is a summary of their remarks.

Criticism

One of the first to comment on the APL approach was H. Acland in 1976. He questioned the University of Texas researchers' assumption that life skills can be measured by a competency test. He argued that: (1) people do not face the same problems, and (2) there are a variety of ways of solving a problem. Acland cites several APL items to show how hard it is to find life-skills problems that all people actually encounter in their lives.

One item Acland analyzed asked people to use a mock airline schedule to select a flight so that they could make an appointment in another city at a particular time. According to the APL study, 30 percent failed to choose the correct answer. As Acland notes, the APL developers apparently assumed that the airline schedule item measured some general life-skill or competence. Nevertheless, Acland observes:

> *The case can be made that there are skills which apply to checking airline schedules which are not necessary for other tasks. In that case it is pertinent to ask how many people use airlines. In September 1974, only one person in two (55 percent) had ever flown (Gallup, 1974). This makes the "bad" result look a lot less depressing; it (the APL finding) now seems to reflect differences in the tasks people face rather than differences in their problem solving skills.*

To illustrate his point regarding different ways in which people solve the same problem, Acland cites results on an APL item indicating that only 39 percent of the population could correctly complete calculations for a partially completed Federal income tax form 1040 when provided with the pertinent information. In contrast, he cites Internal Revenue Service data indicating that only about 6 percent of individual returns actually contain arithmatic errors of any kind.

The two points raised by Acland are essentially questions of measurement validity—ones that are particularly crucial to any effort to assess life-skills competence. First, it cannot be inferred that a person lacks a particular skill simply because he or she is unable to demonstrate it in a testing situation. Second, just because someone lacks a particular skill does not mean that he or she will be unable to deal with a problem in a particular real life situation. People ask for help.

Along these same lines B. R. Heller et al (1978) criticizes APL Study Director, Norvell Northcutt's statement that, "A shocking 34% of the (national assessment) sample felt that police had the authority to detain a suspect for as long as a week without bringing charges against him (not recognizing) the illegality of the situation." Because of the way the items and options are worded, Heller wrote, their is no correct answer: "Under no circumstances" is not correct according to the Constitution of the United State which allows for the one exception if there is a state of martial law. This example is not meant to be lighthearted nor erudite. We include it as an illustration of two problems that are common to the items. First, as an example of poor item construction and/or second, of an elitist value system that suggests that most people need not know "the full story."

Heller also examines the USOE finding that "12 percent of the sample felt that permission to have public meetings should not be given to certain kinds of groups; e.g., 'radicals' and 'troublemakers.'" Clearly, the wording of this item asks the respondent for an opinion rather than the constitutional guarantee; the proportion of "incorrect" responses may not be as surprising when interpreted in this manner.

The Texas researchers defined success in society strictly in terms of income, education and occupational status. Yet, Haney and David show that APL items apparently discriminate far more effectively in terms of people's ethnicity, age and city size, than they did in terms of respondent's income.

Please turn to page 15

—Why I Resigned from ETS—

Following are a few excerpts from Chuck Stone's keynote address to The Committee's, NJEA's, NJPTA's, and NJPIRG's October 1981 conference in Saddlebrook, N.J.

It is a pleasure to be here among so many old friends and new acquaintances. To be honored on the same program with Dr. Banesh Hoffmann, the guiding genius who pointed the way for all of us, is a distinction I shall treasure for the rest of my life.

Over 80 million American voters may go to the polls next month to choose a president and yet not one of them has one word to say about the composition, relevance or fairness of ETS tests which are required today to enter college, law school, graduate schools, business and medicine, in some cities and states, the teaching profession and several occupations and professions such as real estate brokers, automobile mechanics and policemen.

ETS has such a stranglehold on American education because its tests are great validators, the prestigious authenticators, the mandatory legitimizers of higher education. ETS can raise test fees — and it does so periodically — merely to increase what it euphemistically calls excess income over expenses but it translates as profits. ETS is a business corporation with the same major goal as General Motors or any multinational group — make money and may the consumer be damned.

As an organization, ETS has its hands deep into the cookie jar of federal grants awarded because of very strong ETS contacts as a result of an interchanging relationship between federal and private employees. Federal officials know that taking good care of ETS in awarding contracts can result in a good job during a change in administration or a cutback in department appropriations.

The year before I left, 1971, ETS had successfully bid on a project known as Task B-6 — named after the outmoded bomber. Its title was "A Study of the Special Service Programs in Higher Education for Disadvantaged Students" and would assess existing programs for disadvantaged students and identify the elements and characteristics of successful programs in colleges and universities with all-minority and racially mixed student populations for research comparisons.

In the RFP (Request for Proposal) from the Office of Education, I was listed thusly: "Mr. Stone would serve as a kind of 'senior advisor in residence', devoting 10 percent time over the course of the project."

When I resigned in February, 1972, Task B-6 was woefully behind its schedule and overspending its budget. I subsequently received a letter on March 16, 1972 from an ETS executive which read — mind you, now I have resigned — "You may not agree, but I suspect that it is critical for you to retain the title and responsibility of co-director. The trick then (and it would have been a trick) will be to find ways that you can continue to shape the project and the final report without infringement on your new responsibilities."

The letter goes on to state that two of the project's top overseeing officials in the Office of Education had been informed of my resignation and that they would probably want a letter from me that I would not leak any information and a personal letter that my resignation "is in no way prompted by the project."

It wasn't. I resigned because I discovered that a small group of people at ETS led by its executive vice president, Robert J. Solomon, are far more interested in making a profit than educating America's children.

Last year ETS mobilized one of the most expensive lobbying efforts in history to defeat testing reform legislation, a cost which will eventually result in increased test fees. Due to the brilliant and exuberant leadership of the Committee for Fair and Open Testing and many other individuals gathered here today, New York State enacted America's first comprehensive "Truth-in-Testing" legislation. The enactment of this legislation shows that a coalition exists to make the testing corporations accountable to the consumers who pay taxes to keep it afloat. Our task today is to develop strategies that will make this testing reform movement of ours strong enough to drag ETS, kicking and screaming, into the twentieth century.

Once again, it is a pleasure to be here today — thank you.

Chuck Stone, *ETS' former Director of Minority Affairs, is senior Editor of The Philadelphia Daily News.*

Yum-Yum

ETS Conference Center has a new Executive Chef, Brian Eager. According to *The ETS Examiner*, Eager "has prepared cuisine at fine hotels and restaurants on three continents and cooked for such leading figures . . . as Gerald Ford, David Rockefeller, Frank Sinatra, and Jackie Gleason." Gleason once gave him a $250 tip.

Eager supervises a staff of 12, including five cooks, some of whom were trained at the Culinary Institute in New York. "I will concentrate on providing really exquisite food here," Eager says. "I want the Conference Center to have the reputation of being the best restaurant in the Princeton area."

> **In many communities around the country, the quality of education is being measured by the scores children achieve on standardized tests. Education thus falls under the jurisdiction of the language of statistics, and it is a fact that many schools are now designing their programs almost solely for the purpose of increasing their students' mean test scores. Here, it is slightly unfair for me to repeat the joke about the statistician who drowned while trying to wade across a river with an average depth of two feet. The fault is not with statisticians, whose special language is a remarkably useful instrument for uncovering abstract facts. The fault is with those educators who have fallen under its spell and have allowed their purposes to be subverted by the seductions of precise measurement.**
>
> ***—Neil Postman***
> ***Crazy Talk, Stupid Talk***

Continued from page 16

- worried by the fact that no one knew how ETS scored or equated its tests or what the company did with the information they had on over 70 million (today 85 million) Americans.
- worried by evidence indicating that the multiple-choice test format penalized deep and creative students. (See *The Tyranny of Testing* by Banesh Hoffmann, The Crowell-Collier Press, 1962).
- worried by the fact that research including some done by ETS, has shown that coaching improves black students' test scores the most—yet black students are probably the least able to attend coaching schools—schools that can cost as much as $300.
- worried by the fact that in 1979, the American College Testing Program asked five health experts to examine their 126 item multiple-choice health care knowledge proficiency test and eliminate the obviously incorrect answers. ACT researchers planned to estimate the difficulty of their test by the number of the possibly plausible answers that remained. To ACT's chagrin, each expert eliminated as being obviously incorrect between 9 and 26 of the answers ACT stated were correct. (See Fall 1979 *Testing Digest*.)
- worried by ETS' past practices of sending one score to students while sending universities different scores that were derived from secret statistical manipulations that had the effect of reducing minority and lower socio-economic status students' chances for admission. (See *Reign of ETS* by Allan Nairn and Ralph Nader, p. 246)
- worried by the fact that an ETS employee leaked a "confidential" ETS report dated August 14, 1978 outlining the non-profit company's plan to *increase* by $450,000 ETS expenditures to improve "Public and Professional Attitudes Towards Testing and ETS."
- worried by the fact that just one word, or more accurately, one number from a test publisher can unalterably affect a person's life, yet that person must trust ETS that the test is fair and is scored correctly.

How Truth-in-Testing is Working in New York

After taking a recent Scholastic Aptitude Test (SAT), a Pittsford, New York high school student wrote to ETS pointing out that one of the test's reading comprehension passages that dealt with how migrating birds navigate was incorrect.

"Reading false statements does tend to jar one's senses," he said, "possibly causing errors to be made in answering the questions."

Another student identified a discrepancy on ETS' October 21, 1980 Preliminary Scholastic Aptitude Test (PSAT). He proved that the answer ETS selected as correct was inferior to another choice. ETS acknowledged their error and decided to count both their initial solution and the actual correct answer (see back cover).

Michael Simon, a recent graduate of Columbia University's Masters Program in International Affairs who took the LSAT last summer, wrote to ETS that their computer had incorrectly marked a "no answer" in five places where he had entered a reply.

Such increased scrutiny of admissions tests is one result of New York State's Truth-in-Testing law, which went into effect January 1, 1980, and gives college and graduate school applicants the right, after scores are reported, to obtain a copy of the questions asked and their answer sheets.

Test publishers, who uniformly testified against the New York bill, have reacted differently now that the measure has become law. Eighteen months ago ETS Vice President for Law Programs, Thomas O. White, wrote an internal memo to ETS President William Turnbull (Fall 1979 *Testing Digest*) stating that since Truth-in-Testing had become law, publishers should do their utmost to comply with it. Today, White is president of a new non-ETS affiliated organization, the Law School Admissions Service (LSAS) of Newtown, Pennsylvania, which administers several programs previously run by ETS (including the LSAT).

65% Response Rate

The Law School Admission Council (LSAC) has extended the benefits of New York's law to students throughout the United States. On the May 1980 Law School Admission Test (LSAT) over 65% of the test takers requested a copy of their questions and answer sheets.

One of the reasons that LSAT test takers have taken advantage of the law is that the LSAC has clearly informed students about their new rights. Test takers can obtain a copy of the test and their answer sheets by just checking a box on their original registration form.

LSAS President White told *The New York Times* (November 23, 1980) that as a result of the New York bill: "People can see more of the elements of a process that is critical to their career decisions, and they can correct any mistakes."

Law School Admissions Council Director Bruce I. Zimmer, in his testimony before the New Jersey legislature on February 11, 1981, explained that:

> *Test disclosure promotes confidence in the integrity of the test instrument and the testing process. It is symbolic of our public accountability. Indeed, openness helps us in our quality control efforts. Consequently, our Board has decided that, effective in the 1981-82 testing year, disclosure of test forms, answer sheets and credited answers will become a "built-in" and automatic part of our testing program for all candidates. This result comes first because it will be cost effective for us to send test copies to all test takers rather than to handle tens of thousands of orders—but also because test disclosure, if it is technically feasible, appears to be a service our constituencies support and are willing to pay for.*

Please turn to page 12

Committee Co-Director John Weiss drafted America's first federal Truth-in-Testing bill for Congressman Michael Harrington in 1977.

Truth-in-Testing . . .

Continued from page 11

Despite the high request rate for the LSAT, the College Entrance Examination Board (CEEB), the sponsors of the Scholastic Aptitude Test (SAT), maintains that the law gives students rights they do not want. In a November 23, 1980 interview with *The New York Times*, CEEB President George Hanford stated that the percentage of New York students who have so far been willing to pay $4.65 to obtain a copy of their SAT (approximately 5 percent) confirms this. The Board has also imposed a $1.75 surcharge for all New York test-takers.

There is a simple reason for the high school students' low disclosure request rates. ETS and CEEB officials have gone out of their way *not* to inform students about the law. CEEB President Hanford told *The New York Times*, "Why should we go out of our way to push something we don't think is useful?"

Another reason the SAT response rate has been low is that instead of just checking a box on their initial registration form, SAT test takers must fill out a special form to obtain a copy of the test. In several instances this form has not been included with the regular registration form. These special forms are needed because the College Board only allows New York test-takers access to their test questions.

Sponsors of the Graduate Record Examination (GRE) and the Graduate Management Admission Test (GMAT) have joined LSAC in voluntarily making disclosure national policy. National disclosure allows publishers to spread the cost of making new test questions over a much larger number of test takers. An ETS official told *The New York Times* (11/23/80) that were the College Board to adopt a similar policy the additional cost for each SAT examination could be kept to about 30 cents, but "the view was to punish New York State."

In February, 1980, six weeks after the New York bill became law, New York State Republican Senator Kenneth P. LaValle told the following story to illustrate his frustrations with ETS and CEEB officials: "In the early 1960's when Congress enacted auto emission regulations, both General Motors and Toyota hired 1,000 new employees. Toyota hired 1,000 new lawyers." (See Winter 1981 *Testing Digest* for list of lobbyists ETS has hired.)

"Test Quality UP"

ETS President William Turnbull recently told his senior staff that if publishers stand firm, the testing reform movement will wither. "If I read my tea leaves right," he concluded, "the tide of public opinion may be turning."

In spite of Turnbull's call to stand firm, several ETS officials have publicly endorsed the measure. In October 1980, ETS Vice President for consumer Affairs Jean Brittel told a group of students at Harvard University, "Truth-in-Testing legislation has not harmed the quality of our tests — in fact, test quality probably has gone up."

And ETS Senior Vice President for Testing, E. Belvin Williams told *The New York Times* education editor Edward Fisk that New York's Truth-in-Testing bill has prodded his company into "a healthy re-examination" of the organization's internal procedures. Williams concluded that due to Truth-in-Testing "I can say with absolute certainty that the quality did not go down on any of our tests. To the contrary it is better than ever."

Cost of Administering Truth-in-Testing legislation

In FY 1980 the New York State Education Department incorporated the cost of administering the Truth-in-Testing law into their general overhead budget. The state's principle expenses were to file and store the data provided by the test publishers. In FY 1981 the State Legislature made no special allocation to cover the costs of the law.

How Many Tests Are Still Administered in New York

Three days after the New York State legislature enacted Truth-in-Testing legislation, test publishers announced that twenty of the twenty-six tests covered by the new law would no longer be administered in New York State. Most of the publishers who withdrew their tests did not testify at the Legislature's May 9, 1979 hearing on the bill. After a February 1980 hearing, the New York legislature adopted several amendments (outlined in the appendix) and test publishers reinstated fifteen of the previously withdrawn tests. Today 97% of the tests covered by the bill are administered in New York.

What is to be done?

Unfortunately, there is no simple answer to our current over-reliance on norm-referenced, multiple-choice tests. The process through which reforms will be implemented, however, should be analyzed in terms of an evolutionary process.

The first step in this process is to provide teachers, parents, researchers, students and legislators with access to the information they need to make intelligent decisions about how America's students and workers should evaluated.

ETS is currently promoting the findings of a Gallup Poll they commissioned which found that 81 percent of the American public is opposed to government regulation of standardized testing. Truth-in-Testing, however, is not regulation. No boards are set up. No requirements are mandated. All that test publishers must do is provide students and researchers access to the data they need to ensure that tests have been correctly scored and accurately reflect what they are supposed to be measuring.

As Patrick Rooney, President of the Golden Rule Insurance Company and member of the State of Indiana's Republican Governing Council testified in New York, "If this piece of legislation . . . dealt with the insurance industry or with other profit making ventures, there would be no question at all of its passage."

A complete unabridged copy of John Weiss' testimony is available from The Committee, Room 1000, 5 Beekman Street, New York, N.Y. 10038 for $10.00; $5.00 without appendices.

Starting with this issue, **The Testing Digest** *will feature items from tests disclosed under New York's Truth-in-Testing Law. The first test we shall examine is the* **Law School Admissions Test***—a test which is required by every accredited law school admissions department. The following LSAT items focus on issues relating to civil rights and labor union activities. These questions' subject matter could adversely affect certain test takers concentration—and scores.*

civil rights items

3. Before he was assassinated by his brother in 1828, King Shaka built a Zulu empire extending over hundreds of thousands of square miles and that contained some two million inhabitants.

   (A) extending over hundreds of thousands of square miles and that contained
   (B) extensive over hundreds of thousands of square miles, and it contained
   (C) that extended over hundreds of thousands of square miles and containing
   (D) that extended over hundreds of thousands of square miles and contained
   (E) which was extending over thousands of square miles and containing

6. Afrikaans is the language of the ruling party in South
A
Africa and of the Afrikaners, whose votes maintain the
B C D
status quo. No error
E

7. The Supreme Court ruled that it is not inherently unconstitutional for a white suburb to refuse to change zoning rules which practical effect was to block construction of racially integrated housing.

   (A) which practical effect was to block
   (B) which practical effect were to block
   (C) whose practical effect was to block
   (D) of which the practical effects were blocking
   (E) whose practical effects were blocked by

12. Do you think that our university ought to go on discriminating against disadvantaged students by continuing its current admissions policies?
In terms of its logical features, the question above most closely resembles which of the following?

   (A) Do you think that whatever people do is right?
   (B) Should your neighbor stop beating his wife?
   (C) Do you think children should be taught to believe in the devil?
   (D) Should force be used to prevent a person from committing suicide?
   (E) Does power corrupt people and absolute power corrupt them absolutely?

13. McKinney speaks reverently of Du Bois the scholar
A
and dismisses as a bit of elderly foolishness Du Bois'
B
joining the Communist Party at the age of 93. No error
C D E

This item is the only question on the past three LSAT's that featured a famous black person.

union items

4. Dorothy Day, the founder of the Catholic Worker
A
Movement, has been jailed eight times, most recent as
B C D
an illegal picketer for Cesar Chavez's United Farm
D
Workers in 1973. No error
E

11. The building trades unions appear willing to make
A
wages and benefit concessions in order to become
B C D
more competitive with nonunion labor. No error
D E

13. Real wages began to rise long before unions became
A
powerful, and the level of real wages in various
A B
countries bear no relation to the strength of the union
B C D
movement in those countries. No error
D E

25. In the period since 1957 when wages, salaries and
fringe benefits climbed to the highest levels in history
A
absenteeism resulting from real or fancied illnesses
B
have been increasing at an average annual rate of 2.8
C D
per cent. No error
E

—How To Order Tests—

As a result of New York State's Truth-in-Testing Law, copies of actual standardized tests are now available to the general public.

●Scholastic Aptitude Test (SAT)—The four SATs disclosed in New York State from administration in the January-June 1980 period are available by writing to College Board Publication Orders, Box 2815, Princeton, N.J. 08541. Request a copy of *Four SATs* with $3.00 per copy prepayments.

●Graduate Record Examinations (GRE)—GREs may be ordered by using the order form in the Bulletin, which gives dates the tests will be available, or by letter to Graduate Record Examinations, Box 9[illegible]-A, Princeton, N.J. 08541 for $2.00, prepaid. Also available in the same way—by order form or letter—are the two GREs disclosed in the January-June 1980 period, now published together for $3.00 prepaid.

●Law School Admission Test (LSAT)—May be ordered by writing to the Law School Admission Service, Box 2000, Newtown, PA 18940. Send $4.00 prepaid.

●Graduate Management Admission Test (GMAT)—"By policy of the graduate Management Admission Council disclosed tests are not available to noncandidates." (There are sample questions in the Bulletin; a full test is in the Guide to Management Education, available through bookstores at $5.00, or by order form in Bulletin or letter to ETS for $4.00; Graduate Management Admission Test, Educational Testing Service, Box 966-R, Princeton, N.J. 08541.)

——Late Breaking News——

(March 26, 1981) Faced with Truth-in-Testing bills in 26 states and Congress plus convincing evidence that opening up the testing process will improve test quality — see pages 14-16 — College Board Trustees voted today to instruct ETS to partially extend the Truth-in-Testing concept nationwide. Starting immediately high school students outside New York State will be allowed to see their Scholastic Aptitude Tests as well as their answer sheets and a list of correct answers.

Despite the fact that College Board officials recently testified against Truth-in-Testing (March 23 in Massachusetts and March 24 in Florida), *The Times* reported that there was little opposition among College Board Trustees to make disclosure the nationwide policy.

ETS/CEEB EARNS $2 MILLION

The List Business

Reacting to legal action brought by The Committee in a Brooklyn, N.Y. federal court (see Fall 1979 and Winter 1981 *Testing Digests*) ETS/CEEB have started informing high school students that the CEEB owned, ETS operated Student Search Service (SSS) provides "colleges and governmental scholarship agencies" with information about students' race, ethnic background and family income. The SSS obtains their lists from the forms students fill out when they take the one of the 14 CEEB sponsored, ETS administered college entrance examinations.

Last year SSS sold lists containing over 22 million names, earned over $2.7 million in revenue and $2,300,000 in "non-profits" *(excess income over expenditures)*.

SSS still does not inform students that the U.S. military is considered to be a governmental scholarship agency.

"White-Only" Lists

The March 9, 1981 *New York Magazine* reported that last year several colleges used the SSS to purchase mailing lists containing the names and addresses of just white SAT test takers.

* * *

In 1978 4.3 million Americans turned 18. This year the figure will be 4.1 million. By 1992 it will be less than 3.2 million. As one admissions official at a formerly selective college (who requested to remain anonymous) recently lamented: "We're no longer an admissions office, we're a recruiting department."

ETS Caught With Its Integers Down a 2nd Time

A second high school student has successfully challenged an officially correct answer on a standardized mathematics test used in college admissions, in this case the Scholastic Aptitude Test.

As a result, The Educational Testing Service announced on March 23, 1981 that it will upgrade 20,000 New York high school students' test scores.

New York Education officials predicted that 40 to 50 additional students will be awarded $1,000 Regents Scholarships because the error was discovered.

"It Didn't Make Sense"

The student who pointed out the faulty question, 17-year-old Michael Galligan, a senior at Clarkstown South High School in New City, New York told *The New York Times* that he wrote to ETS because Question #16 on his SAT "didn't make sense to me [because] it had two answers."

The question asked:

> *"Which row in the list above contains both the square of an integer and the cube of a different integer?"*
>
> *(A) 7,2,5,4,6*
> *(B) 3,8,6,9,7*
> *(C) 5,4,3,8,2*
> *(D) 9,5,7,3,6*
> *(E) 5,6,3,7,4*

The correct answer was (B), since 9 is the square of 3 and 8 is the cube of 2.

Answer (C) was included as a possible answer for students who recognized that 4 was the square of 2 and that 8 was the cube of 2, but it was marked wrong because the question specified that the integers be "different."

Mr. Galligan marked the correct answer, (B), but at the time he also noted that (C) would also be correct because 4 is also the square of minus 2, which is a different integer from 2. Because he had noticed that there were two answers, he exercised his option to see the test questions and his answers.

ETS officials told *The Times* that Galligan's discovery was a direct consequence of New York's "Truth-in-Testing" law, which requires testmakers to provide students with copies of the test and their answer sheets 30 days after their scores are reported.

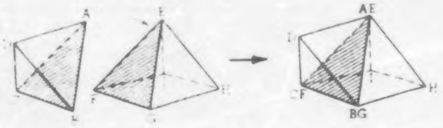
The error did not affect students outside New York State because, unlike sponsors of the Law School Aptitude Test and other admissions tools, the College Board decided not to make "disclosure" its national policy.

SAT's Not Required By 28 Minnesota Colleges

Twenty-eight of Minnesota's 31 colleges and universities allow students to submit Preliminary Scholastic Aptitude Test (PSAT) in lieu of Scholastic Aptitude Test (SAT). The three colleges that still mandate the SAT are small religious institutions. The institutions that accept PSAT scores include Carlton, The University of Minnesota and the entire state college system. The PSAT costs students $3.75. The SAT costs $9.25.

—Solution to PSAT Question #44—

Mr. Lowen proved that answer *(A) five* was correct because faces ABC and EGH and faces ACD and EFI form parallelograms when the solids are placed together (see below).



The Committee For Fair & Open Testing, Massachusetts PTA, Massachusetts Public Interest Research Group, Massachusetts Teachers Association and The National Conference of Black Lawyers are jointly sponsoring an information and ideas sharing conference on:

Evaluation Reform: The Next Five Years

When:
Saturday, May 2, 1981
Where:
Northeastern Law School,
Boston, Massachusetts

For more information contact:
John Weiss
Room 1000, 5 Beekman Street
New York, NY 10038
(212) 349-6460

APL Study

Continued from page 9

For example, the data indicated that all (100%) of the field tested items showed significant associations with years of school, 72% with occupational prestige, and 43% with income. Yet a high proportion of the item results were associated with the test-takers ethnicity (95% of the times), age (86%) and city size (99%).

Haney and David criticize the circular logic used by the University of Texas researchers. They note that "after defining the three APL levels without much explanation, the APL investigators turn around and interpret the results as showing that certain adults are 'functionally incompetent.' This conclusion is, we believe, arbitrary — a product of the study's failure to distinguish between 'incompetence' and low levels of success, both of which were determined by the limited criteria of education, income and occupational status."

They continue: "Even if we put aside doubts about the existence of a general construct of adult competence, the much publicized finding that 20 percent of American adults are 'functionally incompetent' on the basis of the design, conduct and reporting of the APL study is altogether untenable."

Conclusion

The APL study is a classic example of faulty research resulting in naive acceptance and extensive promotion and dissemination, without an in-depth examination of the underlying assumptions. The lesson to be learned from the APL study is that it is dangerous to lose sight of the actual test questions when interpreting test results and formulating conclusions.

Haney and David in assessing the effect of the APL program conclude: "It has helped to broaden conceptions of how adult basic education can be organized and has enriched the body of curriculum materials available to educators." Yet, "it has served to promote a questionably narrow approach to the assessment of adult competency."

Lawrence Kronick, a junior at Connecticut College, is spending a semester working for The Committee.

CHANGE JOBS

Change the World

Every year *Community Jobs* lists over 2,000 job-openings in community and social-change work nationwide. And every month, our special features help you get the job done. If you are a college student looking for an internship, a disgruntled worker looking for meaningful work, or an activist wanting to keep informed about what is really happening at the grassroots — you'll want to subscribe to *Community Jobs*.

SPECIAL OFFER! Get your first issue of *Community Jobs* **Absolutely Free**. Just fill out this no-risk subscription coupon today. We'll bill you later ($8.88 for 9 more issues — 30% off the cover price!) If you don't wish to subscribe, write "cancel" on the bill and keep your first issue free — no obligation!

NAME
ADDRESS CITY
STATE ZIP

1704 R Street, N.W. Washington, D.C. 20009
Published monthly by The Youth Project, a non-profit organization

COMMUNITY JOBS

Federal Government Stops PACE Testing

On February 24, 1981 a group of minority job applicants withdrew their lawsuit against the U.S. Government's Office of Personnel Management when the federal agency accepted their demands that job applicants would no longer have to pass the five-year-old *Professional and Administrative Career Exam (PACE)*. The Federal Government used the PACE test to narrow down the number of applicants for 118 management-level government jobs, ranging from tax law specialist to museum curator.

Between 1975 and 1980 PACE screened out almost all minority applicants. In January 1978, for example, fewer than 1 percent of blacks and only 2.6 percent of the hispanics who took the exam scored high enough to be hired. The success rate for whites was 13.2 percent.

The PACE test will be replaced by between 50 and 60 new exams that will be more closely tied to the specific skills needed in each job category.

Center for the Study of
Education and Politics
The Testing Digest
Room 1000
5 Beekman Street
New York, N.Y. 10038

Non-Profit Org.
U.S. Postage
Paid
New York, N.Y.
Permit No. 7287

EXCERPTS FROM THE TESTIMONY
of
JOHN GORDON WEISS
before the
MASSACHUSETTS STATE LEGISLATURE
EDUCATION COMMITTEE

March 23, 1981

Why New York Passed Truth-in-Testing

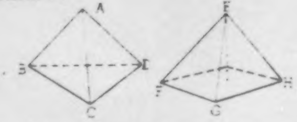
New York enacted Truth-in-Testing legislation because New York PTA, New York Educators Association, New York Personnel and Guidance Association, NAACP, Puerto Rican Legal Defense Fund, New York State Consumer Protection Board, New York State English Council, National Conference of Black Lawyers, New York's Attorney General's Office, and dozens of research scientists, including three former ETS employees, joined the student directed New York Public Interest Research Group (N.Y. PIRG) in enthusiastically supporting the measure.

These individual and organizations supported the bill because they were—

- worried about test publishers' repeated claims that they rarely, if ever, make processing mistakes, when errors affecting at least 450,000 students' scores have occurred in the past few years. (The next issue of *The Testing Digest* will document these errors.)
- worried that The National Conference of Black Lawyers reported that black law school applicants who had comparable grades (± .1 G.P.A.), identical majors, and attended the same undergraduate institutions as their white counterparts, scored over 109 points lower on the LSAT (See Fall 1980 *Testing Digest*)
- worried that ETS awarded students precise "aptitude" scores, when their own studies indicated that there is a 10% chance that a students score will fluctuate over 50 points in either direction if he retakes the test. This fluctuation is important because on most ETS tests a 500 is the 50th percentile while a 600 is the 84th percentile.

(continued on page 11)

Question 44



In pyramids ABCD and EFGHI shown above all faces except base FGHI are equilateral triangles of equal size. If face ABC were placed on face EFG so that the verticies of the triangles coincide, how many exposed faces would the resulting solid have?

(A) Five (B) Six (C) Seven (D) Eight (E) Nine

PSAT October 1980

ETS and the panel of 16 college professors the testing firm hired to review their math tests both agreed that on the above item choice *(C) seven* was correct. Daniel Lowen, a 17 year-old Cocoa Beach Florida High School student, convinced the testing company that answer *(A) five* was correct (see p. 15 for solution).

ETS Senior Vice President for Testing Programs told *The New York Times (3/17/81 pg. 1)* that the question was "a lousy item." He continued, that in the future new safeguards would be added to assure test quality.

PSAT scores are used to award millions of dollars for scholarships and for admissions to Minnesota's colleges and universities. Because the discrepancy was identified, thousands of students' test scores were upgraded by between one and two percentage points—points that could prove the difference between them winning and losing one of the 4,700 financial scholarships and/or 50,000 "letters of commendation" the multi-million dollar National Merit Scholarship Program annually awards students with high PSAT scores.

The PSAT Mr. Lowen took was composed of old Scholastic Aptitude Test (SAT) questions. ETS' Public Relations Director, Robert Moulthroup said, that in the past about 80,000 high school students SAT scores were too low because they selected choice A and were marked wrong.

The October 1980 PSAT was the first PSAT ever to be released to the public.

Appendix D—LSAT questions released as a result of Truth-in Testing legislation

12 Do you think that our university ought to go on discriminating against disadvantaged students by continuing its current admissions policies?

In terms of its logical features, the question above most closely resembles which of the following?

(A) Do you think that whatever people do is right?
(B) Should your neighbor stop beating his wife?
(C) Do you think children should be taught to believe in the devil?
(D) Should force be used to prevent a person from committing suicide?
(E) Does power corrupt people and absolute power corrupt them absolutely?

Questions 3-4 refer to the following passage

A servant who was roasting a stork for his master was prevailed upon by his sweetheart to cut off one of its legs for her to eat. When the bird was brought to the table, the master asked what had become of the other leg. The man answered that storks never had more than one leg. The master, very angry but determined to render his servant speechless before he punished him, took the servant the next day to the fields where they saw storks, each standing on one leg. The servant turned triumphantly to the master, but the master shouted, and the birds put down their other legs and flew away. "Ah, sir," said the servant, "you did not shout to the stork at dinner yesterday. If you had, he too would have shown his other leg."

PRINCIPLE 3

An assault is a threat made intentionally in words or gesture by one person against another to inflict bodily injury by force. It must appear to the intended victim that the aggressor has both the intent and the apparent ability to harm him, so that the person so threatened is put in reasonable fear of immediate bodily harm.

14. As Sally walked home one evening, she noticed that she was being followed by a man she did not know. When she started to run, he ran too, without saying anything. She stumbled, fell and broke her leg. When it was apparent that she was hurt, her pursuer silently turned and ran away. He was caught and was identified as Jones. In a suit by Sally against Jones for assault, she will

(A) win because she fell and broke her leg
(B) win because Jones ran after her
(C) lose because Jones turned and ran away
(D) lose because Jones did not say anything to her

13 Real wages began to rise long before unions became
A
powerful and the level of real wages in various countries
(A) B
bear no relation to the strength of the union movement
C D
in those countries. No error
(D) E

25 In the period since 1957 when wages, salaries, and
fringe benefits climbed to the highest levels in history,
A
absenteeism resulting from real or fancied illnesses have
B C
been increasing at an average annual rate of 2.8 per cent.
(C) D
No error
E

11. The building trades unions appear willing to make
A
wages and benefit concessions in order to
B C
become more competitive with nonunion labor.
D
No error
E

8. A mural painted by Bill Walker and others, "Wall of Dignity" faces a rubble-strewn lot in Detroit's East Side slums, where it presents to residents a history of the black man that begins in ancient Egypt.

7. Undoubtedly because of the tense situation, the school officials rapidly instituted many changes to make the curriculum, the faculty, and the teaching materials more appropriate for the students of the Spanish-speaking community.

7. The Supreme Court ruled in June, 1976, that states
may first try defendants on serious criminal
A
charges without a jury if it later provides the
(A) B
defendants with an opportunity for a new trial
C
before a jury. No error
D E

7. The Supreme Court ruled that it is not inherently unconstitutional for a white suburb to refuse to change zoning rules which practical effect was to block construction of racially integrated housing.

(A) which practical effect was to block
(B) which practical effects were to block
(C) whose practical effect was to block
(D) of which the practical effects were blocking
(E) whose practical effects were blocked by

4. Dorothy Day, the founder of the Catholic Worker
A
Movement, has been jailed eight times,
B
most recent as an illegal picketer for Cesar
C D
Chavez's United Farm Workers in 1973.

No error
E

13. McKinney speaks reverently of Du Bois the scholar
A
and dismisses as a bit of elderly foolishness
B
DuBois's joining the Communist Party at the
C D
age of 93. No error
(D) E

22. The events of September, 1971, in San Francisco
A
illustrate not only the tensions that typically
B
surround desegregation but also the idiosyncratic
C
circumstances that make each city's experience
(C)
with the issue more unique. No error
D E

6. Afrikaans is the language of the ruling party in
A
South Africa and of the Afrikaners, whose votes
B C
maintain the status quo. No error
D E

33. Lengthy unemployment often suggests to those who
A B C
cannot find jobs that they live in America that
(C) D
ignores the needs and aspirations of the common

3. Much of the clamor against the film's sentimental distortion of Jenny Korn's life comes from critics who, inconsistently, raised no protest against the sentimental distortion of Annette Meyer's fiction in its film version.

In which of the following ways might the critics referred to above argue that they were not inconsistent?

- I. There is an important distinction between distorting the story of a person's life and distorting a work of fiction.
- II. Protest and disapproval are acceptable ways for audiences to effect change in entertainment.
- III. Sentimentality is not a distortion but is an expression of genuine feeling.

(A) I only
(B) III only
(C) I and II only
(D) II and III only
(E) I, II, and III

LSAT
June 1980

1

4. How many real men use He-Man Cologne? Well, we've been checking the locker rooms here at Mammoth Stadium. The best pro football teams in the country are with us today. And nine out of ten of these top athletes have He-Man Cologne in their lockers. Shouldn't you be using it too?

It is likely that the appeal of the advertisement above would be weakened LEAST by evidence that

(A) He-Man Cologne manufacturers had given complimentary bottles of cologne to the players just before the game at Mammoth Stadium
(B) people generally believe a man who uses cologne cannot be a "real man" no matter what evidence exists to the contrary
(C) the person who wrote the advertisement neither uses He-Man Cologne nor fits the stereotype of the "real man"
(D) most people do not consider football players to be "real men"
(E) He-Man Cologne is used by more women than men

20. Many cultures do not share our assumptions that women are innately passive and emotional, that their roles must be domestic, and that they cannot function effectively in positions of authority.

30. According to Mary Ellmann, feminine passivity is closely related to black apathy. In each case, having restricted the participation of the group, society finds that inactivity is an innate group characteristic.

38. Susanne Miles has and always will be a person who values her professional reputation above financial gain.

5

17. Cartoons in movies or on television often contains a hero or heroine who perform amazing feats in order to help unfortunate victims.

(A) contains a hero or heroine who perform
(B) contain a hero or heroine who perform
(C) contain a hero or heroine who performs
(D) contains a hero or heroine that perform
(E) contain a hero or heroine which performs

6

[illegible] Feminism Form YLS2-2

Feminist: Opponents of the Equal Rights Amendment claimed that the amendment would deprive women of society's traditional protections. By suggesting that women must be protected by the larger society, these opponents thereby implied that women cannot protect themselves. Those who wish to "protect" women should, by eliminating society's special protections, extend to women an equal opportunity of self-protection.

News commentator: But the average woman is physically weaker than the average man; therefore, only women need and should have the special protection of society.

21. Which of the following could contribute to a sound and relevant argument against the news commentator's claim?

- I. Many of society's traditional protections have nothing to do with physical harm.
- II. Providing society's protection only to women will leave physically weak men unprotected.
- III. A specific difference between average individuals in two classes of people does not necessitate treating those two classes differently.

(A) I only
(B) II only
(C) III only
(D) I and II only
(E) I, II, and III

22. In her argument, the feminist presupposes which of the following?

- I. The main function of society is to protect its members.
- II. All men and women are equally capable of protecting themselves.
- III. Those who wish to "protect" women have the power to extend to women the opportunity of self-protection.

(A) I only
(B) II only
(C) III only
(D) I and II only
(E) I, II, and III

23. Of the following, which expresses the most sound and relevant criticism of the feminist's main point?

(A) Self-protection and protection by the larger society are not incompatible.
(B) The original motives leading to laws to protect women were admirable ones.
(C) All people have the right to life, liberty, and the pursuit of happiness.
(D) Equal rights imply equal responsibilities.
(E) Men as well as women need society's protection.

9. The women enrolled in the judo course in order to learn ways to protect oneself from an experienced instructor.

(A) learn ways to protect oneself from an experienced instructor
(B) learn ways to protect herself from an experienced instructor
(C) learn ways of protecting themselves from an experienced instructor
(D) learn, from an experienced instructor, ways to protect themselves
(E) learn, from an experienced instructor, in what way to protect oneself

Case I

In 1965, Boutin discovered that Chisholm, the owner of a certain farm, had left the state because Chisholm was being sought by the police. Believing that he was entitled to do so, Boutin moved into the farmhouse in that year and acted as though he owned the house. He stayed there most nights, occasionally went away for the weekend, and took two weeks vacation each year. He used the garage and mowed the lawn around the house but did nothing with the balance of the property. The balance had originally been used for crops. Boutin discovered in 1976 that Chisholm had finally been arrested in 1971 and convicted of a crime and was in jail. In 1973, while Boutin was on a short vacation, Perrier used the farmhouse for one day. She caused damage to one of the rooms by starting a fire. The room has since been kept locked by Boutin, who had used it before the fire but no longer did thereafter. In 1974, Lorenz, a neighbor, carelessly started a fire which caused the destruction of the garage. Lorenz did not have Boutin's permission to be on the premises. At the time of the fire, Boutin had been using the garage but thereafter did not use it. Boutin did not repair the structures damaged by fire.

---

37. Unknown to both Malcolm and his family, Malcolm suffered from a terminal and incurable cancer. Malcolm applied to XYZ Insurance Company for a life insurance policy which did not require an examination by a doctor. One of the questions on the application asked whether the applicant was suffering from cancer. Malcolm answered in the negative. Malcolm died two years after the policy on his life was issued, and the premiums were paid in full. When an autopsy was performed, XYZ learned that Malcolm's death had been caused by an advanced, malignant growth on the lungs. XYZ refused to pay Malcolm's wife Norma, the beneficiary of the policy. Norma sued for payment of the benefit. Held, for Norma.

The narrowest principle that reasonably explains this result, and is not inconsistent with the ruling given in the preceding case, is

(A) A contract for life insurance takes effect as soon as the first premium has been paid, and the obligations undertaken in the contract cannot thereafter be invalidated.
(B) A contract cannot be voided by a party to a contract for misrepresentation because of a false statement if the person making the statement, either the other party or his representative, did not know and had no reason to believe that the statement was false.
(C) A party who fails to investigate the truth of statements on which a contract is based cannot later claim that the contract is invalid because of the falsity of those statements.
(D) A contract cannot be voided unless the parties have each been wronged by the other by reason of the falsity of statements on which the contract was based or by breach of promises made in the contract.

3. Before he was assassinated by his brother in 1828, King Shaka built a Zulu empire extending over hundreds of thousands of square miles and that contained some two million inhabitants.

(A) extending over hundreds of thousands of square miles and that contained
(B) extensive over hundreds of thousands of square miles, and it contained
(C) that extended over hundreds of thousands of square miles and containing
(D) that extended over hundreds of thousands of square miles and contained
(E) which was extending over thousands of square miles and containing

11. Kruger supports the Internal Security Act that provides for the indefinite detention without trial of any person whom the police regard to be security risks.

(A) whom the police regard to be security risks
(B) who the police regard as security risks
(C) whom the police regard as a security risk
(D) who is regarded as security risks by the police
(E) the police have regarded to be a security risk

13. McKinney $\underset{A}{\underline{\text{speaks reverently of}}}$ Du Bois the scholar and $\underset{B}{\underline{\text{dismisses as}}}$ a bit of elderly foolishness Du Bois's $\underset{C}{\underline{\text{joining the Communist Party}}}$ $\underset{D}{\underline{\text{at the}}}$ $\underset{(D)}{\underline{\text{age of 93}}}$. $\underset{E}{\underline{\text{No error}}}$

23. Carla was hurrying down the sidewalk toward a subway station on her way to the airport when her way was blocked by Lennox, a tall young man. Lennox said that everyone who passed by there had to pay toll and demanded twenty-five cents. Carla was angered, refused to pay, turned and walked all the way around the long block to her station. As a result, she missed the plane she was going to take and had to take a train to another city, where she had an appointment for an employment interview. The appointment had to be rescheduled and Carla lost two days' pay from her current job because of the delay. Carla sued Lennox for false imprisonment and claimed damages for her loss of pay. Held, for Lennox.

The narrowest principle that reasonably explains this result, and is not inconsistent with the ruling given in the preceding case, is

(A) False imprisonment cannot occur in an open and public place.
(B) Restriction of a person's movement, if the restriction results in delaying but not in preventing a person from executing his plan, does not constitute false imprisonment.
(C) Restraint of movement is not false imprisonment if an alternative way of departure is left open.
(D) Damages for false imprisonment may not be recovered for delays occasioned by the imprisonment.

THE EDUCATIONAL TESTING ACT OF 1981

WEDNESDAY, JULY 22, 1981

House of Representatives, Subcommittee on Elementary, Secondary, and Vocational Education, Jointly With the Subcommittee on Postsecondary Education, Committee on Education and Labor,

Washington, D. C.

The subcommittee met, pursuant to notice, at 9:30 a.m., in room 2175, Rayburn House Office Building, Hon. Carl D. Perkins presiding.

Members present: Representatives Perkins, Weiss, Simon, Erdahl, Bailey, and Erlenborn.

Staff members present: John F. Jennings, counsel; William A. Blakey, counsel; Betsy Brand, minority legislative associate; Richard DiEugenio, minority legislative associate.

Chairman **Perkins**. The subcommittee will be in order.

A quorum is present.

The Subcommittee on Elementary, Secondary, and Vocational Education and the Subcommittee on Postsecondary Education will continue its hearings that we started yesterday on H.R. 1662, a bill introduced by Mr. Weiss and other Members of the Congress, a bill to require certain information to be provided to individuals who take standardized educational admissions tests and for other purposes.

Chairman **Perkins**. We are glad to welcome here this morning a panel of experienced test-takers and testing experts, Carolyn Bennett, a student from Johns Hopkins University; Michael Galligan, graduate, Clarkstown South High School, New City, N.Y.; Dr. Banesh Hoffman, professor emeritus, Math and Physics Department, Brooklyn College; and Dr. Michael D. Priddy.

All of you come around and get around the table.

We will hear from you first, Miss Bennett.

Yes, Congressman Weiss?

Mr. **Weiss**. We have a very outstanding group of panelists today, and I especially want to take note of the fact that Michael Galligan, who is sitting at the table, got up at 4 o'clock this morning in order to get to New York City and catch the shuttle and get to Washington on time.

Chairman **Perkins**. Very energetic.

Mr. **Erdahl**. I would be quite concerned because anybody with that ambition is a potential opponent down the road. That is what you have to do oftentimes.

Chairman. **Perkins**. Go ahead, Carolyn.

178

STATEMENT OF CAROLYN BENNETT, STUDENT, JOHNS HOPKINS UNIVERSITY

Ms. **Bennett**. Good morning, Mr. Chairman and members of the committee. Thank you for allowing me to speak very briefly this morning.

My name is Carolyn Bennett. I grew up as the dependent daughter of an Air Force pathologist. I recently graduated from Johns Hopkins University with a bachelor's degree in natural sciences and will be attending Johns Hopkins Medical School in the fall.

Several summers ago I worked as a summer staff member at Educational Testing Service in Princeton, N.J. My duties included reviewing items for the SAT to insure that any racially or sexually sensitive words or phrases were removed before the items were printed. I also worked on several task forces to provide a college student's perspective on testing issues.

I have also served on an informal advisory group whose function is to exchange information between ETS and some of the groups it affects. Since then I have worked a variety of summer jobs as a medical secretary.

I am here this morning to talk about my experiences with standardized testing.

My first encounter with a nationally administered standardized test was in 1976 when I took the PSAT. At that time I was a junior at a Department of Defense high school in Wiesbaden, West Germany.

The school was average academically. Accelerated classes such as AP English and AP biology were not offered. Though not certain of medicine as a career choice at that time, I was certain that I wanted to attend a 4-year college.

Taking the PSAT helped me out in a number of ways: It introduced me to standardized test format. It started me thinking seriously about what colleges and degree programs were most suited for me. It served to identify me as a candidate for a national merit scholarship for outstanding Negro students which helped pay for my first year of college. And, most importantly, it let me know where I stood in comparison to other college-bound high school students.

This was particularly important as I was tackling college decisions from overseas. There was a small percentage of college-bound students in my graduating class and very little academic interaction with students from other schools. I took the PSAT and knew that I could compete for a spot in a selective college. It was a positive, motivating experience with standardized testing, and not an unusual one.

I took the SAT the following year, in 1977. I did not take a coaching course. I went into the exam armed only with my 4 years of public high school course work, and I walked away thinking that the answer sheet I turned in was a fair representation of my mathemetical and verbal skills. In all honesty, I am not a test-taking wizard. I struggled with a lot of items and found others simple.

The scores I received a few weeks later made me feel good about myself and my academic abilities. I did not receive 700's, but I thought that my scores would give me a crack at a good premedical program, and they did. The SAT did not hinder me academically.

As I took the test, I did not feel as though I was falling into a test trap or that I was being manipulated in some way. I did not find items that offended me as a black woman. The items were fair game to evaluate me as a high school student. I benefitted from taking the SAT. It, too, was a postive test-taking experience.

In my third year as an undergraduate at Johns Hopkins University in Baltimore, I began making plans to apply to medical school. I worked hard in college and received good grades but was not at the very top of my class. Knowing too well how competitive medical school admissions procedures are, I relied very heavily on the MCAT to show me and to show the schools considering me for admission just how my academic strengths stacked up against those of other students.

True, I dreaded taking it and sweated through it, but it was worthwhile. Without it I would not have made the same decisions about what schools to apply to, and I would not have received the same favorable consideration I received from Johns Hopkins Medical School.

The experiences I have just described are not newsmakers. Up to now, there has not been much interest in a string of stories about students who do not despise standardized and use test data objectively. I owe the fact that I am going to the medical school of my choice because standardized tests identified me to the school I am headed for as being capable of doing the work; this was the biggest benefit of all.

The single point I would like to leave with you this morning is that I am not alone. I am not the only college student to have benefitted from standardized testing. My personal circumstances are not too unusual to be shared in some part by thousands of other students, from both public and private high schools, attending public and private colleges and universities and working toward all kinds of professional degrees.

They would agree with me that standardized testing is the means to broader, not narrower, educational opportunities and that standardized tests are critical, fair and accurate tools to use in academic decisionmaking and that they open, not shut, doors.

I would be happy to answer any questions you may have.

Thank you.

Mr. SIMON. Thank you.

Mr. Galligan?

STATEMENT OF MICHAEL GALLIGAN, GRADUATE, CLARKSTOWN SOUTH HIGH SCHOOL, NEW CITY, N.Y.

Mr. GALLIGAN. Good morning, Mr. Chairman and members of the subcommittee.

I would first like to thank Congressman Weiss for inviting me here this morning. It is an honor for me to be here.

In the past couple of months, I have been called upon to examine my feelings about standardized testing very closely. I have also had the opportunity to hear the opinions of other students whose futures will be affected by the scores they will receive on these tests.

I, therefore, hope to speak for a large population of students, that their opinions and feelings may be considered in the important issues which are decided here.

In October 1980, as I was taking the College Board SAT examination, I noticed that to question No. 16 of section 2, two correct answers were listed as choices. This was in contradiction with the directions which indicated that there was one and only one correct answer for each question.

As the truth-in-testing law is in effect in New York State, I was able to receive a copy of the SAT exam, the answer key, and my graded test paper. Upon reviewing the question, I confirmed that there were two correct answers listed and only one was being counted as correct.

I then wrote to the College Board and informed them of the apparent mistake. In March 1981, I was notified by the Educational Testing Service that I was correct and that both answers would be counted as being correct. As a result, the grades of almost 20,000 students who had chosen the alternative answer were raised 10 to 20 points.

Also, an additional 50 students were awarded Regents scholarships as a result of their raised scores. Unfortunately, students outside of New York, where the truth-in-testing law is not in effect, have not had their scores adjusted.

Are not the students in those other States just as correct as the students in New York?

I support the proposal to extend the concept of the truth-in-testing law nationwide. Any test should be based on a mutual agreement between tester and test-taker that there exists a body of knowledge which mankind agrees upon as being true.

It is perfectly possible for a young person to recognize truth by way of insight, even without the formalized education to know it. It is also conceivable that a person with more education but less insight would not see what the student was able to recognize. Twice this has been proven.

Truth is what men live by. To protect, praise, and preserve it is a most holy endeavor. To compromise it is to insult the way we live and degrade the very thing students are told to search for. I know the reason that I wrote away was because I love mathematics. When I saw the mistake, I felt that a terrible injustice had been done. I had nothing to gain from challenging the question but peace of mind.

I feel that if we deny a student the right to see his test, we are breaking a trust with students and ourselves. This country has never been one to deny a person of rights. I think a right that all people have and which is one no person would want to live without is the right to know the truth.

I think I speak for all students in America when I say the truth-in-testing law is a step in the right direction. The promotion of such a concept, as this bill would do, is a positively righteous action.

Furthermore, the Educational Testing Act would assist the student in educating himself by allowing him to recognize his mistakes and, in so doing, learn more about himself and how he functions in testing situations.

While the bill proposes some necessary rights for test-takers, I have to express my feeling that it misses some of the major faults of the Standardized Testing System as it is now. I think the Educational Testing Service should feel a sense of obligation to inform the students of what they will be tested on.

Students as well as universities should be aware of how this test should be interpreted and how accurate or inaccurate it may be. This should be followed up by studies of how the tests are being used and how much weight is given to them.

Are they a primary criterion or are they used to highlight other information about a student? The testing organizations themselves must take responsibility for how the tests are used. If they don't, no one else will.

No test is able to perfectly measure that which it is designed to measure. There is always a margin for error. This is true of all tests and other indicators by which students are evaluated. Looking at all the indicators and other personality characteristics is the only way to get an accurate picture of a person's accomplishments and potential.

I have seen a person's dream ruined because he scored 20 points too low on the SAT exam. I think when an exam is given more consideration than factors such as motivation, insight, virtue, and determination, there is something wrong with our system of evaluation.

To provide equal opportunity for the individual and for the strength and growth of our country, our testing system must be improved. It should be consistent with the values of our society. If standardized tests are to be used as widely as they are now, they should be available to everyone regardless of geographic and economic status.

The users of such tests, namely the educational institutions which require them, should also share in the responsibility of making the tests available. It may involve financial contributions to cover additional expenses. Also, they should be keenly aware of the limitations of such a test.

By limiting our interpretation of others to a number on a piece of paper, we are losing those qualities about a person which truly serve the needs of our country; qualities such as honesty, integrity, charity, compassion, and wisdom, all of which have marked the great men of our country.

There is no way to test for this type of gift; there is no way to measure one's ability to perform, to contribute, or to care. There is no way to see if a person will succeed or fail at something except to give him a chance. There are those who have little chance, who have to strive to overcome barriers of culture, poverty, and location to try to educate themselves.

It is not possible to provide everyone with a very fine education, but it is possible to give everyone the facts about what they should learn for a particular test so that those who do have the motivation can educate themselves and also seek it at their schools where it should be taught.

Our fine government should be an example for the efficiency and integrity of a system of checks and balances.

Mr. **Weiss**. I think we will go to the questioning of the two people who just testified since they are the test-taker part of our witnesses this morning. After that we will go to Dr. Priddy and perhaps Dr. Hoffman will be hereby at that time as well.

Carolyn, I think I perhaps did you an injustice in not asking you how long it took you to get here.

Ms. **Bennett**. Twenty minutes.

Mr. **Weiss**. I wonder if each of you could give us some idea, not just of your own personal experience, but of your involvement with other students' attitudes toward the tests, as to their impression of whether, in fact, there is sufficient information given about them, whether, in fact, they have sufficient opportunity to study for them, whatever.

You are both obviously outstanding students and you may not have some of the same problems others might.

Carolyn?

Ms. **Bennett**. Thank you. My feeling is that my peers and the students that I have spoken to, both about the SAT and MCAT, is generally positive. I emphasize that for the MCAT as well. It is better if I discuss the MCAT in particular.

Mr. **Weiss**. When did you take it?

Ms. **Bennett**. In the spring of 1980. It is critical to the pre-med population at Hopkins to do well on the exam and everybody was coaching and practicing and they want to see as many of the items as they can. It is competitive and takes 6½ hours, but generally it is a positive feeling. The feelings of the people that I took the exam with was that when they came out of it it was a fair representation of the work that they had one in college and it was going to be a good index as to how they were going to perform in medical school.

Mr. **Weiss**. Did most of the people who took that test study for it? Did they attend coaching or preparation school?

Ms. **Bennett**. About a fifth of the medical class took a coaching course.

Mr. **Weiss**. That was the new MCAT?

Ms. **Bennett**. That is right.

Mr. **Weiss**. In 1979, I believe, the medical college testing group decided that they really needed a new test. They were unhappy with the old test, I guess from the point of view of reliability, and they decided that they would try a new series of types of questions, a new area of limitation as to where they would ask the questions from.

Even the testmakers themselves do not contend that they have always given tests which they have been fully satisfied with.

Ms. **Bennett**. I believe that the students that were taking it when I was taking it believe that the material on the exam was appropriate for the purpose.

Mr. **Weiss**. My understanding is that MCAT is limited to the second year college equivalent biology course level.

Ms. **Bennett**. Yes.

Mr. **Weiss**. Everybody knew going into that test what the general area would be that would be covered?

Ms. **Bennett**. Level of difficulty?

Mr. **Weiss**. Well, the subject area that would be covered.

Ms. **Bennett**. Yes.

Mr. **Weiss**. That is sort of different from the SAT, though.

Ms. **Bennett**. True.

Mr. **Weiss**. Did you have a clear idea going into the SAT what areas would be covered?

Ms. **Bennett**. Yes; I did, but because of a different reason, from the PSAT.

Mr. **Weiss**. Michael, what was your experience?

Mr. **Galligan**. Going in to take the SAT exam, I had very little idea as to what specifically the format would be. I had not taken the PSAT so I had no idea in that respect, nor did I receive information from my school as to what the format was.

However, I was fortunate enough to have a fine educational background, so that I was able to handle what was presented to me OK. A particular experience which I would like to tell of is when I went to visit the Jackie Robinson Foundation Youth Convention in Columbia University a couple of months ago.

The students there also did not have the information, the appropriate information, as to what the test would actually test, as to what the format would be and what the necessary levels of math and verbal skills had to be attained so that one could actually do the exercises of the test and display his knowledge in that manner to make the test valid.

Nor did they have the necessary skills in many cases, OK, so I feel that the responsibility really cannot lie with the school. If you are going to say that there has to be complete equality for this test which is designed to measure aptitude so you have to have complete equality as much as you can, then each student has to be personally given the information and it is necessary that all the information is gotten to each student so that he is fully aware of what the tests will be, because in many cases the students did not know what they had to attain educationally nor were they taught it at their schools.

I can't speak for myself in that manner. I have to speak for them because that is where the fault is. I was fortunate enough to go to a very fine school but there are a lot of other people in this world who aren't.

Mr. **Weiss**. How do you feel about the policy which ETS and that which the College Entrance Examination Board people have now adopted, of disclosing upon request the answer sheets together with the questions, in I think five out of the seven major administrations of the test?

How do you feel, Carolyn?

Ms. **Bennett**. I support it.

Mr. **Weiss**. Expand on it if you will. Why do you think that that is a good idea?

Ms. **Bennett**. I think it is important to give students an opportunity to see a copy of their exam but I also think, from what little I know about testing, that it is important to keep some of the exam secure.

Five out of seven is a good decision for ETS to have made in that case. If I was taking the SAT and I wanted a copy of my exam, I would take it at one of the disclosed administrations, and not at one of the secure ones.

Do you see what I mean?

Mr. WEISS. Yes. Why do you think it is worthwhile for students to be able to get their answers back?

Ms. BENNETT. Well, in all honesty, I don't know if I would request a copy of my SAT. On the MCAT, that I would but not on the SAT.

Mr. WEISS. Why would you request it in the MCAT?

Ms. BENNETT. Level of difficulty of the exam, particularly on the quantitative analysis questions where the answer takes quite a while, and I would be able to check levels of difficulty, but I wouldn't feel the need to on the SAT.

Mr. WEISS. That has to do with some extent with your sense of confidence about the subject matter in the two tests, is that right?

Ms. BENNETT. That is right.

Mr. WEISS. Michael, how do you feel about the disclosure policy?

Mr. GALLIGAN. I support it for two reasons: First, as I said, there is an educational benefit in seeing a test which you took and what you got wrong; OK. For just the sake of knowledge, knowing what you got wrong and learning that.

Second, for seeing how you test, which is a very important skill. The second reason is that no one has a corner on truth. Truth is something which everybody has to contribute to, young and old, regardless of age or religion or anything else. Everybody should be able to see what they did so that they are convinced that either it was wrong or right and graded accordingly.

My experience with the SAT testifies to that. I was dissatisfied with the particular question from a personal standpoint, and I think it is very important that a person is able to satisfy that lingering feeling about a particular question.

It is very important to make the truth consistent with the person. He has to understand what is right and what is wrong. A very good way to do that is just to allow people to see the tests.

Mr. WEISS. Carolyn, did you make application to get your MCAT results and were you able to get them?

Ms. BENNETT. No; they are not available.

Mr. WEISS. Thank you very much.

Mr. Bailey?

Mr. BAILEY. I don't have any questions.

Mr. SIMON. Let me add my commendation to both witnesses.

I think it is a great thing to have student witnesses, meaning no disrespect to other witnesses, but we too often hear from professionals in the field and not often enough from people like the two of you. Thank you.

Mr. WEISS. Again, my appreciation and that of the subcommittees to both of you for taking time out of your leisure or worktime to be with us today. Thank you so much.

We will proceed with the rest of the panelists, so if Patrick Shields, Dr. Schafer, Dr. Kasteen, Mr. Sjogren and Dr. Robinson will join Dr. Priddy at the witness table, we will have one large panel and Dr. Priddy will be the lead witness.

You will have to move the microphones back and forth in front of you.

Dr. Priddy, why don't you proceed.

Do you have a prepared statement that you submitted to us?

Mr. PRIDDY. No, sir.

Mr. **Weiss**. I should tell all of the witnesses that if you do have prepared statements we will enter those statements in their entirety in the record without objection.

You may read it, extrapolate or excerpt from it, handle it whatever way you like.

If you wish to add additional material to the testimony you give here the record will remain open for a 10-day period and if you submit it to the committee during that time. We will include other additional material in the record as well.

Dr. Priddy?

STATEMENT OF MICHAEL D. PRIDDY, DIRECTOR OF RESEARCH, PLANNING, AND EVALUATION, GUILFORD COUNTY SCHOOL SYSTEM, GREENSBORO, N.C., ACCOMPANIED BY MARIAN EPSTEIN, EXPERT IN TEST DEVELOPMENT

Mr. **Priddy**. Thank you, Mr. Weiss. I would like to thank the other members of the subcommittees for the opportunity to speak before you.

I am Michael Priddy, and I live in Greensboro, N.C.

I work in a school system in that area. It is in the Piedmont section of North Carolina, industrial service area with a strong agricultural base and a county of about 315,000 people.

I work for the Guilford County School System, 25,000 students, sixth largest in North Carolina, and made up of 44 schools.

Before I proceed with some detail about my work and some experiences, I would also like to make the Chair aware that I am accompanied by Dr. Marian Epstein, nationally recognized expert in test development, and she can respond to specific questions by types of tests administered by ETS.

My work involves planning for instructional and administrative purposes, research, that is the conduct or the oversight of it, evaluation of curricular programs as well as administrative programs and the standardized testing program.

That program is made up of a State-mandated program that involves grades 1, 2, 3, 6, 9, 11, and is completed by a local component that fills in at some of the other grade levels.

If I may, I will limit my remarks to that area specifically, testing, and as it is related to program evaluation.

The plain thrust of the office and its staff is to serve as an adviser to the superintendent on the board of education, and if you will, functions as a gyroscope for the school system in a sense of attempting to keep us on track and insuring that our resources and efforts are pointed toward things that will improve the curriculum and instruction for the students.

For that reason, we draw on multiple sources but one is standardized test results, and I will share some of those experiences with you, particularly the effect of those results on curriculum instruction.

For years we have been accused of letting others dictate what we do, test developers, book publishers, and it is very easy to fall into that trap because they are expert at what they do and public school people can easily rely on them, but that does not have to be the case and it is not the case in many school systems and one is ours.

We take those results each year and very systematically go through them on a system basis and on a school-by-school basis in an effort to determine where the strengths are and where the weaknesses are.

If I might use one example that many people are currently familiar with, minimum competency testing, which has become a pervasive force in America in the last 4 or 5 years, and I think because of public reaction, that is, a perception that we were not producing people who were minimally literate in reading and math.

That wasn't the case in Guilford County when tests were first administered. We did have a small percentage of students who were identified as needing additional help, but the point is, even with the better students, we could see trends in those test results that indicated that the curriculum and the instructional program in the sixth, seventh, eighth, and ninth grades needed some mending, needed to be defined, refined, and revised.

Why do you see this kind of trend, this trailing off later in the grades and we were looking at the text and curricular guides and finding various resons for it? We were and we are making decisions on the basis of standardized test results, feedback, program evaluations, feedback from committees, and so on.

Then there is the other group that are affected besides teachers, and that is the students, and, of course, that is what we are all about.

We found that using multiple indicators as a way to determine where those students stand is a far more appropriate method for insuring that we are addressing their needs, the individual's needs, as opposed to random processes that vary from year to year.

I would like to stop with that for a moment and move to an earlier part of my career when I was teaching. I chose to teach in an area of North Carolina for 4 years, rural, deprived, made up predominantly of minority populations.

When I first arrived at the school system, I learned many things about it, and most of them were biases and prejudices of the people that were there. One prejudice I formed very early is that students were being encouraged not to strive but to do what others before them had done, and that had been based on experience of guidance counselors and principals and particularly with admission to postsecondary institutions.

Students were being encouraged to accept a lower quality of postsecondary instruction than they could handle and that was often associated with standardized test scores, and things like that. Those schools will not consider you, things like that.

We attempted to alter that view, some colleagues and myself, and began pointing out to students that the test scores are only one of many indicators used by colleges and universities to make decisions, and particularly in North Carolina I can think of two or three schools at that point in history were making sincere efforts to look at grades, community involvement, school involvement, so on, and used in the standardized test scores as an indication of present ability, and the others as independent indicators of potential.

I say that with some confidence, because my wife worked in admissions at one of those schools and I am drawing on her experience rather than mine there.

I think the critical point there, though, is the tests for both achievement and ability were providing some indications of where people stand and those in turn could assist the student and the educators with decisions about where they could, can, and should go.

If I might drop back and catch some of that in another phase, yesterday I had the opportunity to work with a group of people representing various organizations, universities, political world, the State bureaucracy, and so on, and a question which this group has mandated be addressed and that is, what is adequate education in the United States, and how should that be promulgated by the Federal Government.

Congress has asked that position papers be prepared on the subject and submitted by December and we were in fact working on one of those yesterday.

Among other things, we concluded that adequacy must be defined in terms of society as well as the individual and to have some indications of what is adequate, one must have some indicators which can be many different kinds of things.

In particular, regardless of limtiations, because they are in our world, we need some things that allow comparisons across groups and among individuals, and standardized testing for us allows that, particularly things like minimum competency testing. We need to know whether or not students can read, write, and recollect and, if they can't, what are we going to do about it.

Then you get into some questions when you are talking about things like that, whether or not students are adequately prepared to deal with those matters, and I think that is the responsibility of the public schools.

We can do that by informing parents, students, and working with teachers, and so on, and that matter can be addressed very easily.

In summary, I would simply leave you with the idea that when you have a school system like Guilford County that was already performing above national norms when the mandated tests were first administered in North Carolina, that does not leave you in a position of satisfaction.

That leaves you in a position of being able to do more and we just reported to our board last night—our board of education—although I was not there because I came up here, that over a 4-year period scores have continually gone up in all areas across the system, and in every school in all the areas measured.

I think that is what we are all about in education, improvement, and do those kinds of things and to show other people there must be some standard way of doing it.

Mr. WEISS. Thank you, Dr. Priddy.

Mr. Shields?

STATEMENT OF PATRICK SHIELDS, DIRECTOR, EAST HARLEM COLLEGE AND CAREER COUNSELING PROGRAM, NEW YORK, N.Y.

Mr. SHIELDS. Good morning. My name is Patrick Shields, presently the associate director of an organization known as the East Harlem Career Counseling Program.

My past experience in the field of education includes working as an assistant dean in the admissions office at Amherst College, includes teaching literacy up in Massachusetts for the New England Farm Workers Council, as well as running a neighborhood center in Holyoke, Mass.

My masters work is in educational administration at Columbia University, with emphasis in Federal policy, and I am also presently serving on an advisory committee for the College Board for a program they are trying to put together to identify talented Hispanic students throughout the country.

I have been asked to come here because I am from New York State and we have a truth-in-testing law, and so I have worked under it for the last year and I have been asked to speak about that, but it is important to note first where I work and how I work.

I work in East Harlem and the students with whom I work range in age from 14 to 27 in East Harlem and the South Bronx. Those names, of course, ring bells in your ears. They are very impoverished areas. East Harlem has an unemployment rate of around 50 percent, median income between $5,000 and $6,000, depending on the group we are talking about, whether or not it is Hispanic or black.

Those particular aspects of the community, of course, have their effects on the students I work with. The lack of employment opportunities, the welfare dependent lives of so many of the families, affect students greatly.

They are, of course, the ones for whom there are the fewest occupational jobs opportunities available in the city. The educational opportunities available to my students to make up for that, in no way compensate for the bleak occupational outlook.

In New York, 45 percent of the students who enter high school in the ninth grade do not graduate. That is in the city as a whole.

In our local academic comprehensive high school in East Harlem, there are 1,800 students, 71 of whom graduated, 7 of whom graduated with New York State Regents diplomas.

Statistics, of course, can't really talk about what those students have to go through. They don't really represent the reality of the situation there.

Walking into a high school in the morning and seeing the kids out on the steps getting high, or playing basketball over in the park across the way, says a great deal more than those statistics do, yet somehow too many of those students sometimes for our likes overcome those incredible roadblocks and continue to strive for some ill-defined measure of success in their eyes

In other words, they stay in school or they try to get back into school. During the school year, our appointment calendar is filled up 2 and 3 weeks in advance. These are eighth- and ninth-grade students who want to go into high school, students who dropped out and want to get back in. They are, however, in the majority, students who are in the 11th, 12th grades who are faced with the maze-like process of testing and financial aid and admission forms.

It is our job as college counselors to help them through this maze. Unfortunately, the simple fact that they have made that decision to continue their education in no way does away with the obstacles that have been thrown up in front of them in the educational process.

Unfortunately, too many of those obstacles are in the form of standardized tests, and my students perform extremely poorly on those tests. It is not my purpose today to expose the negative side of standardized tests in America. The validity of those tests has been argued back and forth on both sides of the political spectrum.

However, the facts and arguments of those particular arguments concerning standardized tests are germane to the discussion of how my students confront the Scholastic Aptitude Tests.

Standardized tests play an extremely important role in the educational life of the students who come to see me. In such a large system as New York, there are many, many educational options. A student coming out of the elementary school, going on into junior high school, has a number of junior high schools to choose from, and within the junior high schools, there are specialized schools.

Moving into high school, there is a myriad of five options that the student has. New York has the distinction of having some of the very finest high schools in the country—and some of the worst.

The decision on who gets into the very best high schools is made upon a student's performance on one standardized test. That is not true of all the high schools.

There is no question that the guidance counseling in the schools depends a great deal upon the students. In other words, where they are guided to go to high school depends upon their results on standardized reading and math tests. Almost all of the important decisions in a student's schooling have been based all or in large part on the performance of standardized tests, and so, by the time they reach my door, they have come to perceive those tests as having some type of mystical powers.

Most frequently, this is fostered in the form of a negative self-attitude. These tests have proven them to be, in their own eyes, somewhat stupid. Given this mysticism of the test, it is often very difficult for me to prepare a student for the Scholastic Aptitude Test.

This difficulty is compounded because the test is becoming more important every year for my students in terms of their chances of attending the colleges of their choice. This is not because the admission process is getting tougher. There are less students applying to college now. It is because of the role of the test in the financial aid process.

As the Federal financial aid dollar continues to shrink in both real and inflationary terms, the competition for colleges' own campus-based money becomes much more keener, and as a total applicant pool decreases, colleges begin to compete for those students by buying them.

I know of many schools where scholarships are offered regardless of need, any kind of financial need, for three-quarters of tuition for a certain combination of SAT scores, and full scholarships for a certain SAT score combined with a comparably higher rank in class.

Very few of the students ever receive any of those merit-based moneys, so when a student first comes to me for assistance in the college application process and I tell them that they have to take the Scholastic Aptitude Test, I really am faced with a dilemma because a student, the type I work with, almost always appears extremely anxious. Their most frequent response is, "I know I am going to fail it." I see a need to try to lessen that anxiety but, at the same time, it would be unfair of me to minimize the importance of the test.

So, confronted with this when I first began to work at the East Harlem Counseling Program, I spoke to some people in other organizations and decided to establish a seminar that the students could attend for the 3 weeks before the examination where we could begin to deal with some of these problems.

The purpose of the seminar was not to further develop the students' verbal or mathematical abilities; unlike commercial prep courses, we didn't hand out vocabulary lists and algebraic and geometric concepts. We set up a course in such a way students became intimately familiar with the examination, the types of questions asked, the way the questions were phrased, and the amount of time.

The whole purpose of the course was to get the student to a point that they were so familiar with the exam and the process of taking it that they no longer harbored the fears and anxieties that he first encounters.

Mr. Weiss. The bells have rung. There is a vote on the floor of the House, so we will take a 10-minute recess.

[A short recess was taken.]

Mr. Weiss. The committee will resume its hearing.

If you will pick up, we can continue.

Mr. Shields. Thank you.

I was talking about the seminars I had established to help students prepare to take the Scholastic Aptitude Tests. After a few of the seminars, two serious flaws became evident which are relevant in today's discussion. Both of these problems resulted directly from my need to resort to various commercial SAT prep guides in the absence of copies of the actual test.

Given the nature of the seminar, to instill confidence through familiarity, not having real SAT questions in the students' exercises obviously reduced our chances of success. Without actual test questions, it was difficult in most cases and impossible in many to convince the students that they were walking into the exam with accurate knowledge of what they would confront.

The second problem resulting from lack of access to the test concerned my effectiveness as an instructor of the prep course. Commercially published prep books and Educational Testing Service publications on the SAT contain all the different types of questions that are asked on the tests. The commercial texts, however, do not contain questions of consistently high quality.

You cannot compare the text and detect subtle emphasis within certain sections. The truth-in-testing legislation remedied those two problems for me. With actual copies of the SAT to show the students, I have been much more successful in instilling confidence in my students.

I believe it is telling to note attendance has doubled since we began to use the actual SAT in the course. Our preparation has been significantly more successful, and they react very favorably to having the actual copies of the tests themselves. My coaching has improved since I have been able to study the exams and compare the tests.

The enactment of truth-in-testing in New York State has also allowed us to add another dimension to our counseling services vis-a-vis the SAT's. Before the law, if a student came to us after taking the SAT with the feeling that he had done poorly, or if his scores were way out of line with what we had expected, we had no recourse but to sit down with the student and try to remember what he had done. Rarely were we ever successful in finding any problems.

Now, we can request a copy of the examination, the actual corrected test, sit down and begin to analyze where and how the student went wrong. This may afford us the ability to reveal concrete, nonaptitude related reasons for poor performance.

For example, a student may have guessed too often, rushed answers at the end of certain sections in order to finish and not leave blanks and, very importantly for some students, the students with the highest test anxieties, they are able to sit down and review all the questions and answers and realize that they could have performed better on the test, that they were capable of answering questions that they missed.

If their poor performance on the exam had been due to a test anxiety, they can begin to understand and deal with it.

I came here today to speak because of my personal experience with the effects of truth-in-testing legislation in New York State. I have outlined that experience, and showed its overwhelmingly positive effects on my work, with students in East Harlem. I do hope, however, that my discussion today has attested to more than just this experience.

It is my opinion that if the Educational Testing Act of 1981 is enacted and carried out in good faith, it will have its greatest effect on the most educationally disadvantaged of the students who take the SAT. It will be my students' counterparts in the urban and rural ghettos across this country who will be given a much fairer shot at succeeding in the educational process.

Truth-in-testing promises these students the opportunity to confront their attitudes toward the tests directly. As declining enrollments make tests less important for the rich who can afford the full cost of education, the shrinking financial aid dollar and the pernicious practice of no-need, merit-based awards make tests more important to the poor.

The Educational Testing Act of 1981 recognizes that "there is a continuous need to ensure equal access for all Americans to educational opportunities of high quality." It is, however, beyond the scope of both this bill and of the powers of the Federal Government in general to legislate how students are educated on the elementary and secondary levels and how colleges choose to spend their private moneys.

What you can do, given the present situation, is to give students like the ones I work with in East Harlem just a little better shot. You can provide them with one more instrument in their fight to overcome the obstacles inherent in our educational system which have served to keep them down in the past.

I urge you to enact the Educational Testing Act of 1981.

[The prepared statement of Patrick Shields follows:]

PREPARED STATEMENT OF PATRICK M. SHIELDS, ASSOCIATE DIRECTOR, EAST HARLEM COLLEGE, AND CAREER COUNSELING PROGRAM, NEW YORK CITY

Good morning. My name is Patrick M. Shields. I am presently serving as the Associate Director of the East Harlem College and Career Counseling Program in New York City. Ours is a not-for-profit Talent Search agency which works with over one thousand students annually, assisting them to formulate and carry out their educational and career goals. My previous experience in the field of education includes teaching literacy for the New England Farm Workers' Council, directing a neighborhood center in the Puerto Rican ghetto of Holyoke, Massachusetts, and working as an Assistant to the Dean of Admission at Amherst College. My Masters work at Columbia University is in Educational Administration with an emphasis in Federal Policy.

I have been asked to speak here today because of my experience with the Scholastic Aptitude Test after enactment of truth-in-testing legislation in New York State. Our clientele at the East Harlem College and Career Counseling Program(EHCACCP) ranges in age from fourteen to twenty-seven and comes predominantly from the East Harlem and South Bronx areas. These are neighborhoods which are designated by the Mayor's office as poverty areas within a city of manifest poverty. Our most recent statistics on El Barrio, as East Harlem is termed by its residents,show an unemployment rate hovering around fifty percent and a median family income of between five and six thousand dollars. The lack of employment opportunity and the dead-end, welfare-dependent lives of so many of the area's residents has its most chilling effect on the youth of East Harlem. I recently asked a young woman who was in my office advertising a

Yet somehow many, very many of these young people, overcome most of these incredible roadblocks and continue to strive for some ill-defined notion of success. The demand for access to our program is evidence of that fact. During the school year, it takes from two to three weeks to get an after-school appointment. These are eighth and ninth graders who are seeking assistance in choosing high school; they are teen-agers who left the frequently unbearable conditions of their high schools, but now want to return to the educational system after experiencing the frustrations of an already too tight job market without a high school diploma; they are tenth and eleventh graders who need curriculum guidance in their preparation for myriad professions; they are, in the majority, seniors in high school preparing for college and faced with the maze-like process of testing, financial aid, and admissions forms. The EHCACCP is a community agency, not affiliated with any educational institution, and so none of these people have to come to see us. Thus, our clientele is a highly motivated, self-selected group. This does not mean, however, that they are any less educationally or culturally disadvantaged than their counterparts who, for the moment, stay out on the street. Our students have simply reached a point in their lives where they are able to take what they see to be long-range, positive steps. It is our job to help them to take these steps; the fact that they are resolved to succeed in no way lessens the obstacles which they will face throughout the educational process. Unfortunately, too many of these obstacles are thrown up by standardized tests. Our students perform extremely poorly on these tests.

It is not my purpose here today to expose the negative side or prejudicial nature of standardized tests in America. It is a known fact that black children score on the average one standard deviation below their white counterparts on standardized intelligence test. The validity of these tests continues to be argued by scholars on both sides of the political spectrum. The constitutionality of the use of standardized tests to assign public school pupils has been battled out in a number of Federal Courts. Most recently, both the Fifth and Ninth Circuits have ruled some of the use of these tests unconstitutional. These facts and arguments, although not the focus of today's hearing, are germane to a discussion of how minority students confront the Scholastic Aptitude Test. Standardized tests have played an extremely large role in the educational lives of our students in East Harlem. In such a large school system as New York's, the educational options are much more numerous and broader than they are in small systems with only a handful of schools. Moving from elementary school to junior high, students are faced with the options of various specialized schools in addition to various programs within their local school. Moving from junior high into high school, the options are nearly limitless. New York City has the distinction of having some of the very best high schools in the country and some of the very worse. Admission to the very best of the schools is based upon a student's performance on one test. Even in the case of schools where admission is not based on a standardized test, students are often counseled on where to apply or not to apply based upon their performance on standardized reading and math test which they take throughout their school careers. The point here is that almost all of the important decisions in a student's schooling have been made based all or in a large part on their performance on standardized tests. By the time they reach our door, they have come to perceive these tests as possessing some mythical powers. Most frequently, this has fostered a negative self-attitude. These tests have proven them to be, in their own eyes, stupid. Given this mythicism of the test, it is often difficult to help prepare students to take the Scholastic Aptitude Test. This difficulty is compounded by the fact that their performance on the SAT is becoming more important in their chances of attending a college of their choice. Interestingly, but not surprisingly, this is not because the SAT is becoming more instrumental in admission decisions. Rather, SATs are more important now because of their role in the financial aid process. Last year at the private colleges in my state, New York, only fifty-five percent of the accepted applicants received sufficient financial aid. And we are talking about a state with a generous financial aid program of its own. As the Federal financial aid dollar continues to shrink, in both real and inflationary terms, the competition for the colleges' own campus-based aid becomes much keener. As the total applicant pool decreases, colleges compete more for students by literally buying them. I know of schools where scholarships are offered, regardless of need, for three-quarters tuition for a certain combination of SAT scores and class rank, full scholarships are offered to those with better SAT scores and a comparably higher rank. Very few of our students, even our very best who are accepted at the most selective institutions in the country, receive these solely merit-based monies.

And so when a student first comes to me for assistance in the college application process and I inform him that he will need to take the Scholastic Aptitude Test, I am faced with a dilemma. The student is usually extremely anxious about the test. The most frequent response is, "I know I'm going to fail it". As a counselor, I see a need to lessen the student's anxiety, to demythicize the role of the test. At the same time, it would be both unrealistic and unfair to the student if I were to minimize the importance of the SAT. After discussions with my colleagues at the East Harlem College and Career Counseling Program revealed that their students exhibited the same intense anxiety and fear at the mention of the SATs, I established a SAT preparation seminar for our students to deal with these problems. The purpose of the seminar, then, was not to further develop the students' verbal and math abilities. Unlike many commercial SAT prep courses, which are normally much longer than our three week seminar, we do not hand out vocabulary lists and review all of the algebraic and geometric concepts. Nor do we attempt to key the students into any magical test-taking tricks. I simply set up a course in such a way that the students become intimately familiar with the format of the SATs, the type of questions which are asked, the way questions are phrased, and the amount of time alloted for various sections of the test. The sole purpose of the course is to demythicize the SAT, to get the student to the point that he is so familiar with the exam and the process of taking it that he no longer harbors the fear and anxiety that characterized his first encounter with the idea of taking the test. Our hope is that with a more realistic perspective and a more positive attitude, he will be able to perform up to his potential.

After a few of the seminars, two serious flaws became evident which are relevant to today's discussion. Both of these problems resulted directly from my need to resort to various commercial SAT prep guides in the absence of copies of the actual test. Given the nature of the seminar, to instill confidence through familiarity, not having real SAT questions in the students' exercises obviously reduced our chances of success. Students were told that they were taking practice tests which resembled the Scholastic Aptitude Test, but I could not honestly tell them how close the resemblance actually was. Without actual test questions, it was difficult in most cases and impossible in many to convince the students that they were walking into the exam with accurate knowledge of what they would confront. The second problem resulting from lack of access to the test concerned my effectiveness as an instructor of the prep course. Commercially published prep books and Educational Testing Service publications on the SAT contain all the different types of questions that are asked on the tests. The commercial texts, however, do not contain questions of consistently high quality. The ETS publications did not provide a sufficient quantity of questions such that one could begin to detect subtle emphasis within the different sections. For example, there are certain algebraic concepts which are emphasized more on the math section than others. Without access to a number of tests, I could not begin to get the type of grasp of the exam I needed to be an effective teacher.

The truth-in-testing legislation which was enacted into law in New York State remedied these two problems in our seminar on the

SAT. With actual copies of the SAT to show to the students and from which to compile exercises, I have been more successful in instilling confidence in my students. These are young men and women who have grown up surrounded by a plethora of social service agencies offering to help them in some aspect of their lives. They are extremely perceptive of programs such as ours and have learned to distinguish quality assistance from the ineffective services offered by many agencies. I believe that it is telling to note that attendance at the SAT seminars has doubled since we began to use the actual SAT exams in the course. There is no question in my mind that our preparation has been significantly more successful in easing the anxiety in our students. They simply react very favorably to having actual copies of the SAT in their hands. They feel as though they know what they are getting themselves into. Likewise, my coaching has improved since I have been able to study a good number of SATs and begin to compare the tests.

The enactment of truth-in-testing in New York State has also allowed us to add another dimension to our counseling services vis-à-vis the SATs. Before the law, if a student came to us after taking the SAT with the feeling that he had done poorly, or if his scores were way out of line with what we had expected, our only recourse was to have the student sit down and try to remember everything he did or didn't do during the exam. We would then try to figure out where he had gone wrong. Rarely were we successful in uncovering the problem. Now we can request a copy of the student's actual corrected test. We can then proceed to analyze where and how the student went wrong. Not only does this serve to further demythicize the exam, but it may also reveal concrete, non-aptitude related reasons for a poor performance. For example, a student may have guessed too often, or may have rushed answers at the end of each section in order not to leave blanks. Importantly for some students, being able to sit down with the test and review the questions slowly demonstrates to them that they could have performed much better, that they were capable of answering questions that they missed on the exam. Their poor performance may have been, again, due to a test anxiety that they can now begin to understand.

I came here today to speak because of my personal experience with the effects of truth-in-testing legislation in New York State. I have outlined that experience and showed its overwhelmingly positive effects on my work with students in East Harlem. I do hope, however, that my discussion today has attested to more than just this experience. It is my opinion that if the "Educational Testing Act of 1981" is enacted and carried out in good faith that it will have its greatest effect on the most educationally disadvantaged of the students who take the SAT. It will be my students' counterparts in the urban and rural ghettos across this country who will be given a much fairer and better shot at succeeding in the educational process. These are the students who perform most poorly on the SAT and so those who have the most to gain from passage of the legislation. It is these students who are most alienated from the testing process. Truth-in-testing promises these students the opportunity to confront their attitudes toward the tests directly. As declining enrollments make tests less important for the rich who can afford the full cost of education, the shrinking financial aid dollar and the pernicious practice of no-need, merit-based awards make tests more important for the poor. The "Educational Testing Act of 1981" recognizes that "there is a continuous need to ensure equal access for all Americans to educational opportunities of high quality". It is, however, beyond the scope of both this bill and of the powers of the Federal government to legislate how students are educated on the elementary and secondary levels and how colleges choose to spend their private monies. What you can do given the present situation is to give students like the ones I work with in East Harlem just a little better shot. You can provide them with one more instrument in their fight to overcome the obstacles inherent in our educational system which have served to keep them down in the past. I urge you to enact the "Educational Testing Act of 1981".

Mr. **Weiss**. Thank you very much.

Mr. Schafer?

STATEMENT OF WILLIAM D. SCHAFER, PROFESSOR, EDUCATIONAL MEASUREMENT AND STATISTICS, COLLEGE OF EDUCATION, DEPARTMENT OF MEASUREMENT AND STATISTICS, UNIVERSITY OF MARYLAND

Mr. **Schafer**. Mr. Chairman, members of the subcommittee, my name is William Schafer, associate professor of measurement, statistics, and evaluation in education at the University of Maryland.

It gives me great pleasure to be here on behalf of the American Personnel and Guidance Association, the National Professional Association of over 40,000 personnel, guidance, and counseling workers, and the Association for Measurement and Evaluation in Guidance, one of its national divisions.

I appreciate very much this opportunity to express the reservations I have concerning several aspects of the Educational Testing Act of 1981.

I wish to preface my remarks by noting that the Association for Measurement and Evaluation in Guidance has adopted a position with respect to legislation affecting testing for selection in educational and occupational programs.

In that statement, we identify an environment in which tests can contribute most effectively to society and indicate that there is no obvious need for legislation affecting testing, since most test producers already follow the practices we recommend.

We do recognize, though, that others may feel differently regarding the need for legislation on testing, and have developed a series of provisions which we might support as well as a list of proposals which we recommend against. A copy of the statement is appended to my remarks today.

One major concern I have about H.R. 1662 is reflected in my association's statement, which recommends specifically against enactment of legislation which requires disclosure of items following testing. The provisions in sections 5(a) (1)(A) and 5(a) (2)(A) of H.R. 1662 would do just that.

I see little value and much harm in disclosure. The value lies in examinees being able to check their scores and evaluate on a subjective basis the items they took. On occasion, as has happened already, they or others may convince a test producer of the correctness of a nonkeyed answer, thus affecting test scores.

On the surface, this seems to argue strongly that test review by examinees, available from full disclosure, is desirable. However, this benefit is not without qualification.

Let us consider the value of score changes. Competent test use entails the interpretation of a result as indicative of a range of potential values for a score as opposed to a specific point value. An analysis of responses to test items is used as a basis for establishing the range and the extent to which poor items exist is reflected in its size. Thus, a change in scoring for some items on a test is unlikely to move the score range of a given examinee totally outside its original limits.

The harms of disclosure are many. First, assuming the items are to be reused, effects of inequality in access to information about a test for future examinees would probably increase, since the items themselves would be part of that information. Also, the measurement properties of the items would be affected, likely harmed. Thus, disclosure would effectively destroy further usefulness of the items.

On the other hand, if the items are not to be reused, new items would need to be developed for each test administration. This would result in more expensive tests as well as poorer tests, since the best items could be used only once.

Second, without the capability of reusing items, the ability of test producers to equate tests would be impaired, reducing comparability from administration to administration and thus the ability to study trends over time, unless nonscored items were included for equating purposes. While some producers already include nonscored items, the net result is often a poorer test since, in general, the greater the number of items comprising a score, the more accurate it is.

Third, attention of future examinees may be diverted from what could be valuable preparation experiences to merely a review of a large number of previous items. And fourth, since items over certain content areas are likely to be easier to produce than over others, the proportional content of a given test may be altered over time from that which has been judged most appropriate.

The net result of these factors would be harmful to the predictive properties of tests. But some may ask what harm there is in reducing test effectiveness since many other predictors of success are in common use as well.

To address this point, it should be noted that the relationship between tests and other predictors is less than perfect. This means that there exist individuals for whom a low test score is one of the few indicators of a low likelihood of success, and others for whom a high test score is one of the few indicators of a high likelihood of success.

If the impact of tests on admissions processes is reduced, as I think it would be with mandatory disclosure, applicants for whom a high test score is one of the few positive indicators would be the losers, in favor of those who can present other positive indicators but who do not score well on tests.

But is this harmful? Yes, since the other indicators will be used regardless of the weight given tests, or whether or not they are used at all.

Thus, tests are providing additional information, and it is axiomatic that additional information cannot harm, that is, can only improve the performance of a selection process. Can it be wrong to allow all segments of society the full range of opportunities to demonstrate potential?

I should like now to examine several other provisions of H.R. 1662. In section 3(a)(3), relationships between test scores and career success would be required. Actually career success is not easily defined, let alone measured, and is affected by choice among subspecialties offered at postbaccalaureate institutions, which information is not available at the time of application.

Tests are intended to be predictive only of success in studying for a specific career and not of success in a career after studying for it. Therefore, no test within the range of tests specified in this bill in section 2(a)(3), namely "admissions and placement," attempts to predict career success. The only role played by the considered tests in a person's professional life is in selection for admission, and it is on this basis that they should be judged.

Section 3(a)(4)(A) would require it to be determined to what extent test scores improve prediction of grade point average over and above all other information used. But the indicators used by various institutions differ, and with varying success. It seems to me that it is sufficient that a test producer convince a given institution of the value of its product, and I have no objection to this information being made public.

However, justifying the use of any element in a selection process should be the responsibility of the institution using it, not of the supplier of some aspect of the procedure. No test producer can speak for the varied procedures used at all institutions, and surely it cannot be the intention of the bill to impose a mandatory, uniform selection process on all institutions.

In the same section, it is specified that the extent of improvement be expressed as a percentage. This is ambiguous and has been the source of some abuse of statistics. The basic question is: A percentage of what?

The most widely accepted measure of prediction accuracy is a squared correlation coefficient or some variation thereof. Assuming the use of multiple linear regression for prediction, there are two relevant square correlations: With and without the test as a predictor, the former being the larger. Test critics prefer to use the difference between these, which can be interpreted as a percentage. Test supporters tend to divide the difference by the square correlation without the test, which can also be interpreted as a percentage.

I would recommend the latter method, since it reflects the ability of the other indicators to predict grade point average as well as the value of the test, itself. As a minimum, more specific direction is needed. However, I am rather skeptical of the value of either of these indices for a public untrained in statistical methods.

Section 3(a)(4)(B) calls for an examination of the relationship between test score and income level. This invites interpretation of a casual relationship between these variables which is based on questionable assumptions. Moreover, such data are available now to competent researchers who are skilled in its interpretation.

Section 3(a)(4)(C) would require an analysis of the value of test preparation course. Since in most cases the tests we are discussing measure learned abilities, it is reasonable to expect that such programs can improve test scores, though available evidence indicates that short-term experiences do not do so appreciably. It seems to me that it is more the responsibility of these programs to advertise themselves than it is the responsibility of test producers to do so.

Test preparation courses are the province of numerous private tutors, schools, and even books, very few of which report to testing agencies. Separation of the effects of these programs would be a monumental task, but necessary for meaningful interpretation of the data. I should also mention in passing that a test score average is not suitable as the base for a percentage, as called for in this section.

Section 4(a)(1)(A) would require making public studies about or using these tests. This seems to me to carry the potential of discouraging research unless test agencies were also required to provide test data under certain conditions. This area needs more thought, which should be undertaken before enacting a provision such as this one.

I also note that, while subject anonymity is guaranteed, institutional anonymity is not. This would also tend to discourage valuable research.

I would like to close by pointing out that I am not at all against scrutiny of the testing industry by the public. Actually, I feel the testing industry, as a whole, is turning in a creditable performance. But I find H.R. 1662 to be, in general, both harmful and misdirected. I would prefer to see, instead, inquiry into the uses of tests and other indicators of success, as well, made by institutions.

Many of the suggestions in the Association for Measurement and Evaluation in Guidance statement appended here are pointed in this direction.

Thank you.

Mr. **Weiss**. Thank you, Dr. Schafer.

Dr. Casteen?

STATEMENT OF JOHN T. CASTEEN III, DEAN OF ADMISSIONS, UNIVERSITY OF VIRGINIA

Mr. **Casteen**. I am John Casteen, the dean of admissions and associate professor of English at the University of Virginia.

I have come today as a representative of the National Association of College Admissions Counselors, with a mission of sharing with you a description of a research project and planned national forum on a publication having to do with college admissions tests that my association will undertake.

The prepared materials given to you include a prepared statement, which will be a somewhat longer version, as well as a copy of our proposal and, finally, an appended article to which I will refer to in a moment.

Mr. WEISS. Your entire statement will be entered into the record in full.

Mr. CASTEEN. I am a member of the American Personnel and Guidance Association, American Association of Collegiate Registrars and Admissions Office, a trustee to the College Board. The NACAC plan calls for three tests, and the defenses made with regard to it as it relates to the college process, the college admissions process, and not other aspects of testing.

We intend, first of all, after having secured proper financial support, probably from foundations, to commission a group of four investigative papers having to do with specific aspects of the controversy about college admissions tests. In my prepared materials I have described the likely contents of these papers as well as likely authors for them as a way of indicating the scope and the quality of investigation.

The titles proposed at the moment include:

What is the "Truth About Testing?", a paper that would explore many of the issues that have been referred to earlier by members of this panel, a paper having to do with admissions tests and the admissions process, trying to describe with some accuracy the number of different processes used to indicate the rating given, different credentials in the process and experiences students and institutions have had with the process.

Third, a paper having to do with alternative ways to match up students and colleges, a paper exploring specifically other kinds of admissions tests that might be used to provide the kind of ongoing and comparative data that we now get from standardized tests, as well as ways to match up students in a larger sense, to see to it that they do not incur the wasted experience that some students have when they begin.

Fourth, a paper that will embody some recommendations to your committee, entitled "College Admissions Tests and the Public Interest," trying to define on the basis of the first three papers what in large terms ought to be the Federal or congressional stake in college admissions tests, and what national policies perhaps including a document with legislation like that proposed today that might serve college admissions.

After having secured production of those papers, the National ACAC intends to sponsor a national forum on college admissions tests and the admissions process to invite representatives of the major professional and academic organizations that serve this area, representatives of the public, and we would hope to have representatives of your committee attend and take part in a forum that will deliberate the findings of the four papers, make further recommendations that we might sponsor concerning revisions to be made to the papers and forward a document to be the NACAC report to the public in the form of a published book.

We envision having the papers completed by early fall and expect to convene the conference in late fall or early winter and publish the report as soon as possible thereafter with an eye toward providing a timely document for your use.

We want to achieve what we hope will be a balanced analysis of the facts.

In my prepared statement, I said we are apprehensive that the bulk of the material currently available to the public and to your committee is political and partisan in nature, and that the organizations have not devoted the kind of time, energy, money, and concern to the project that we think it deserves, and we intend to do that.

We want to inform the admissions profession about the controversy in admissions testing and to see to it that our organization has practitioners that have the best possible advice.

Third, we do want to pursue alternative ways of matching up students and colleges. We believe other methods are possible. We are conscious that in Great Britain there is a very different kind of admissions test, the discursive test that is represented by level A and O examinations serves the system, and that students progress to college in other parts of the world and, finally, we hope to offer reasoned recommendations to the public and to the Congress concerning the proper public stake that exists in this area of admission to college.

In concluding this overview of my prepared materials, I need to express some areas of personal concern that I recall having first expressed to your committee in 1979 when I came to you as a teacher and dean to express my personal concerns about legislation proposed to you at that time.

Education's problems are at this point quite large. We have problems of funding, serious problems of teacher morale and concerns about the quality of teacher preparation. Ample evidence that the curriculum currently existing in high schools is less than optimal in its service to our students, ample evidence that we are less successful than we ought to be in serving nonmainstream students with access to higher education.

No serious critic of the educational scene at this point blames either declining achievement throughout the system or the continuing inequity of opportunity and achievement that separates nonmainstream, especially racial minority students, from mainstream students or students from relatively poor economic backgrounds from relatively rich economic backgrounds or what appears to be evidence of the diminished overall effectiveness of education during the last 15 years or so, the kind of evidence that turns up in the report of the President's Commission on Foreign Language Study.

I am concerned to note that while your committee is deliberating standardized tests, no serious critic who has looked at these problems blames the test for these conditions. We do have initiatives within education to try to see to it that we face squarely some of the problems that have been evidence.

You have heard one described earlier by my colleague from Guilford County in attempting to decide what schools ought to do. I have attached a newspaper article describing the College Board's Project Equality, which is one of several efforts to address the question of what goes on in the academic courses that students take in high school, whether to prepare for college or life in general.

It seems to me Congress may or may not find sufficient cause to enact H.R. 1662 or analogous legislation, but that it ought to do so only in light of a broader examination of American schools and schooling than has yet occurred, and my concern in the largest sense is not so much the presence of this proposed bill or of earlier bills of the same kind, as it is the absence of initiatives undertaken by the Congress in the direction of solving other problems of which I suspect the testing controversy is merely one symptom.

It seems to me we need to look into the whole question of how schooling works in this country and to separate the core issues having to do with what students learn in schools from what I see as procedural issues having to do with how we measure and report and compare those results in student to student.

Thank you.

[The prepared statement of John Casteen follows:]

Prepared Statement of John T. Casteen, III, Dean of Admissions, University of Virginia

I am John Casteen, Associate Professor of English and Dean of Admissions in the University of Virginia. I came to you in October 1979 to express concern about what I believed would be the adverse effects of passage of H.R. 3564 and H.R. 4949 (as then written) on education in this country. At that time, I spoke in my own behalf as a teacher and dean. Today, I represent the National Association of College Admissions Counselors (National A.C.A.C.), a professional association of some 3000 individual practitioners of college counseling in high schools, colleges, and other settings. My missions are to share with you a report on the progress of the National A.C.A.C.'s task force on college admissions tests and to urge you to delay passing legislation until receiving our report on tests and the college admissions process.

The National A.C.A.C.'s members voted at our fall 1980 meeting to call a national forum on college admissions tests, the admissions process, and other related educational matters, and to provide a report that may serve you and other lawmakers who take an interest in the controversy about tests. To this end, a special task force has produced a comprehensive description of the project, and the National A.C.A.C.'s Executive Director has begun seeking financial support for it. I have provided for the record a copy of this description as it now exists. Our plans for the forum are all but complete at this time, and our leaders are working to recruit the best available talent to write investigative papers for our eventual report.

We plan to commission four investigative reports:

--What Is the Truth about Testing?

--Admissions Tests and the Admissions Process

--Alternative Ways to Match Up Students and Colleges

--College Admissions Tests and the Public Interest

Each investigator will work with an advisory committee to develop a suitable outline and to agree on basic methods and goals. A special National A.C.A.C. task force will oversee the designs of all four investigative papers in order to guarantee that all relevant topics are covered. We expect to receive complete drafts of these four documents by early fall. We can provide draft copies for you at that time.

In late fall or early winter, depending on funding and on the progress of our four papers, the National A.C.A.C. will convene a national forum on college admissions tests. The formal description identifies likely participants. We hope to persuade members of this committee to take part or to attend as observers. Each investigator will present his or her draft report to this forum, and forum members will advise the investigators and their advisory committees on the content, methods, and conclusions that ought properly to appear in the final report. We expect to hold this forum in the Washington, D.C., area and to attract participants who will bring expertise, commitment to thoughtful deliberation, and credibility to it.

Following the forum, the National A.C.A.C. will publish its report on college admissions tests. This report will include the four investigative papers in their final forms, a suitable introduction and perhaps afterword written by the moderator of the forum, and any other documents that may seem appropriate to the National A.C.A.C.'s Executive Board. We will provide copies of the final report for you and for others interested in the issue as soon as possible after the forum. We will offer additional copies for sale to the public at cost.

The National A.C.A.C. has entered into this project only after the most serious deliberation. This investigation marks a new departure for us. Our previous endeavors have addressed the professional concerns of high school counselors and college admissions officers. We have not previously addressed issues of public policy. At the same time, our members and leaders include admissions professionals and others competent to conduct investigations in the field and to offer recommendations that may prove useful to you. By voting to sponsor this investigation, the members of the National A.C.A.C. committed themselves to collect the facts, to seek conclusions based on the best evidence and experience available, and to share this information with you and with the public. Everyone in education probably has some opinion on college admissions tests. We want to get beyond opinions, whether ours' or others', and to offer the substance that has so far not appeared in the controversy about admissions tests.

When I came to you in 1979, I spoke about my concerns about the accountability of our systems of education and especially my sense that Congress might well enhance the quality of life in the U.S. by addressing the problems of our nation's schools and colleges. Since that time, little has occurred to allay my concerns about our schools. In recent months, the Secretary of Education has found it necessary to speak out on what he terms the flabbiness of our public high schools. National news magazines have repeatedly told us that schooling does not work as well as it should for many of our children. James S. Coleman and others have contributed to our awareness that we need to recapture the excellence that our educational system developed (however briefly) in the years between Sputnik and Vietnam. Virtually all commentators have told us that we need to address the core academic skills--reading, writing, arithmetic, foreign languages, and the like. With support from the Ford Foundation, the College Board has begun an investigation of college preparatory curricula that is (from my personal perspective) the most exciting new development in education since the Conant Report. (I am providing a description of this project that I wrote in May for the Richmond News Leader.)

With others in education, I welcome your willingness to explore our problems and to seek remedies. We look forward to working with you to understand what problems exist and to seek solutions that will work.

July 18, 1981

Proposal for the National Association of College Admissions Counselors Forum on College Admissions Tests

1. Background

In both 1979 and 1980, the National A.C.A.C. annual Conference considered motions on the controversy about college admissions tests. In 1979, the National A.C.A.C. elected not to take a stand in the testing controversy. In 1980, the National A.C.A.C. voted by a substantial majority to take this position.

> Resolved: That the National A.C.A.C. go on record as opposing passage of the Weiss Bill as presently stated, or any legislation currently proposed to regulate the design or use of standardized tests in the admissions process or to regulate the process itself; and that National A.C.A.C. now issue an urgent call for a national conference of competent professionals, representatives of appropriate professional and academic organizations, and spokespersons for major test preparation companies to provide a responsible public forum of professional

counselors and admissions officers from which may come creditable independent evaluations of the development, construction, and use of standardized tests.

The effect of this motion is to ask Congress and state legislatures to delay further legislation in the areas of admissions testing and the admissions process pending a comprehensive professional inquiry to be initiated by the National A.C.A.C. In a letter dated March 20, 1981, National A.C.A.C. President Ann P. Fritts and Executive Director Charles A. Marshall communicated the sense of this resolution to the Honorable Carl D. Perkins, Chairman of the Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor of the U.S. House of Representatives. Similar letters went to others who have taken an interest in college admissions tests.

The resolution does not commit the National A.C.A.C. to any prescribed position in the controversy about tests. Rather, it commits the National A.C.A.C. to investigate the charges currently made with regard to college admissions tests, the predictive validity of the credentials used in admissions, and the admissions process itself, and to publish its findings. The members of the National A.C.A.C. believe that the relevant facts about college admissions tests have not been fully examined. As a professional association whose members deal on a daily basis with college admissions tests and the admissions process, the National A.C.A.C. feels responsibility to undertake an appropriate investigation.

This proposal calls for the National A.C.A.C. to commission a group of studies on the various elements of the controversy about college admissions tests, then convene a national forum of competent and scholarly professionals from relevant disciplines to conduct a full inquiry and publish its findings.

Three constituencies stand to gain by a properly prepared report: the general public, whose confidence in the admissions process may have deteriorated in recent years; admissions counselors, who look to the National A.C.A.C. for guidance on professional matters; and federal and state legislators, to whom the National A.C.A.C. ought to present a thorough, competent, and reasonable explanation of the controversy about admissions tests. To date, no balanced investigation of the charges and responses has appeared. The controversy has developed through a series of partisan, polemical documents. The National A.C.A.C. believes that public policy ought to be grounded in more objective and complete evidence than we now have.

2. The Investigative Papers

The National A.C.A.C. inquiry is to move in three stages: advance preparation of four investigative papers, each addressing a discreet topic relevant to the controversy and each prepared by a prominent investigator in an appropriate discipline; convocation of a national forum to receive these four papers, debate the issues, and propose revisions;

publication of the National A.C.A.C. report on the controversy about college admissions tests, this report to include the four investigative papers, any other materials that may seem sufficiently important to the National A.C.A.C. Executive Board, and a general overview of the controversy. The National A.C.A.C. is now seeking foundation support for all phases of this project. The Executive Board believes that the arrangements can be completed by late summer 1981. The National A.C.A.C. is now negotiating with persons qualified to prepare the four investigative papers. Our goal is to convene the national forum in late fall or early winter 1981, and to publish the National A.C.A.C. report as soon as possible thereafter.

The National A.C.A.C. will commission special investigative papers on these topics:

--What Is the Truth about Testing?

--Admissions Tests and the Admissions Process

--Alternative Ways to Match Up Students and Colleges

--College Admissions Tests and the Public Interest

In each instance, the National A.C.A.C. will seek an investigator whose name is known to the community of counseling and admissions professionals, and whose accomplishments prior to this project qualify him or her to render advice useful to Congress and to admissions professionals.

"What Is the Truth in Testing?" will be directed to a general audience, including the public, admissions professionals, and federal and state legislators. Topics to be addressed include: What are college admissions tests supposed to predict? What can such tests determine, and what can they not determine? Within what limits can test score reports be said to be accurate or inaccurate? How faithfully do the makers of college admissions tests tell the public and the admissions profession the facts about their products? To what extent does reliable evidence suggest that admissions professionals use test results properly or improperly?

"Admissions Tests and the Admissions Process" will also be directed to a general audience. Topics to be addressed include the variety of admissions systems used by our colleges and universities and the predictive value of all commonly used admissions credentials. As the investigation may bring to light non-traditional credentials or unusual combinations of credentials that can improve the quality of admissions decisions, the National A.C.A.C. investigators will describe their findings so that the public and the admissions profession may know about alternatives to current practices. With regard to minority students, this investigation should consider whether reliable evidence supports the claim that admissions tests discriminate or support discrimination. These questions are likely to matter: What role have college admissions tests played in opening or closing access to higher education for populations that were excluded before the tests came into widespread use? To what extent have admissions professionals used test score reports properly or improperly in assessing the credentials of non-mainstream populations? In instances of genuine abuse, how can admissions professionals protect applicants from the use of inappropriate selection criteria?

"Alternative Ways to Select Students" will be directed primarily to a more specialized, professional audience, whose expertise in admissions will require that the author be a prominent admissions professional. Topics to be addressed include alternatives to the college admissions tests currently used (no tests, discursive tests, and so on), and their implications for education, alternative ways to fit colleges to students and students to colleges, and the essential strengths and weaknesses of current methods and available alternatives.

"Admissions Tests and the Public Interest" will be directed to a general audience, with special concern for the requirements of lawmakers. Topics to be addressed include an overview of the controversy, suggestions for further or related research, and the public policy implications of college admissions tests and the sundry processes by which students progress from secondary to higher education.

The National A.C.A.C. will designate advisory committees made up of admissions professionals, academics in appropriate disciplines, and distinguished citizens to referee these papers as they are produced. Each author will confer with his or her advisory committee before beginning work and will submit an outline to the National A.C.A.C. task force charged with seeing that the topics detailed above are treated in sufficient breadth and depth. Each author will submit a complete draft to the appropriate committee of referees in sufficient time to permit revision as the author and the committee of referees may agree. The National A.C.A.C. will provide clerical support, copying services, and reasonable support funds for original research in accordance with the principles of good practice that apply to the research community.

The National A.C.A.C. Executive Board envisions an initial report (prior to the national forum) of ca. 150 pages, including the four investigative papers, a suitable introductory statement by the Executive Director of the National A.C.A.C., and perhaps brief notes to be prepared by the designated moderator of the forum. To be produced inexpensively and as a draft, this document will go in advance to forum participants and to federal and state legislators who may request drafts, with notes indicating that the final publication will differ from the pre-publication.

3. The National Forum

In late fall or early winter 1981, the National A.C.A.C. Executive Board will convene a national forum to assist the investigators and advisory committees in revising the drafts of the investigatives papers and to debate the issues of the controversy about college admissions tests. Certain constituencies must be represented in this forum; others may be invited as they express interest in participating and as the

National A.C.A.C. Executive Board comes to believe that they may contribute significantly to the deliberations. The professional leadership of the National A.C.A.C., the American Personnel and Guidance Association, of national pyschometric and psychological associations, and of other academic or professional groups with an interest in college admissions tests will be included, along with recognized individual practitioners in admissions, psychometrics, and related fields. After consulation with the National A.C.A.C.'s eventual financial backers, the National A.C.A.C. Executive Board will decide how to include the College Board, the American College Testing Company, the Educational Testing Service, and perhaps other test makers and test sponsors, as well as their more prominent critics. Commissioners of Education from states where testing legislation has been an issue (New York, California, Louisiana, and so on) will be included if they have conducted independent research into the validity of admissions tests or the practicality of legislation proposed or enacted in their states. Testing's lay critics will be included if they provide academically sound documentation in suppport of their published claims.

Representatives of other constituencies, including the U.S. Department of Education, the National Science Foundation, the American Academy of Science, members of Congress concerned with legislation about tests, and perhaps the National Education Association and American Federation of Teachers, will be invited to take part.

In the public interest, the National A.C.A.C. will seek to enlist prominent participants from the lay public, including citizens who have not previously taken part in the controversy but whose capacity to judge the evidence and to speak clearly on it is well established.

All forum participants will be invited to offer written comments on any aspect of the topic, with the understanding that the National A.C.A.C. Executive Board will exercise sole discretion in determining what documents in addition to the four commissioned papers ought to appear in the final National A.C.A.C. report on the controversyabout college admissions tests.

4. Outcomes

The National A.C.A.C. realizes that this topic is a volatile one, and that this project may matter to groups outside the admissions profession, including lawmakers. The National A.C.A.C. intends to produce a comprehensive report embracing relevant issues that may present themselves in the course of research and debate, and to share this report as widely as possible with concerned citizens. The National A.C.A.C. realizes that testing's critics and friends will have equal motives to "win" the debate; at the same time, the National A.C.A.C. believes that its members, general public, and the nation's students deserve uncommonly high standards of objectivity and professional responsibility. We seek to offer a reasoned report, perhaps offering alternative judgments of the validity of the evidence or the cogency of the arguments. The National A.C.A.C. wants this report to be the definitive word on the "truth" in testing, as well as a durable guide for the immediate future for legislators, students and their parents and teachers, and admissions professionals. We believe that these persons have the right to know what college admissions tests are worth, how they work, and what they can and cannot do for or to an individual student. In addition, this report will treat the integrity of admissions credentials, guidelines of good practice for admissions officers, and perhaps also for admissions test takers and makers, and statements of specific advice to federal and state legislators who must decide whether to regulate college admissions tests. With the seriousness of these purposes in mind, the National A.C.A.C. will enforce throughout this project the strictest standards of professional ethics and academic rigor, for itself and for all participants.

Prepared by John T. Casteen, III
for the National A.C.A.C.
Task Force on Admissions Tests

12 EDITORIAL PAGE The Richmond News Leader Friday, May 22, 1981

GUEST COLUMN:

On the Competencies Students Ought to Have

By John T. Casteen, III

LAST WEEK in St. Louis, about 130 college and school students, teachers, deans and principals, school-board members, and other educators met to hammer out a working proposal for overhaul of the college preparatory curricula offered in our schools. Under the sponsorship of the Ford Foundation and the College Board, the group reviewed recommendations from subgroups that met earlier to assess what kinds of high school backgrounds students need to succeed in college, and came to agreement on two statements. One statement describes the basic competencies (writing, reading, doing mathematics, and so on) that all students need for college. The other describes in general terms the kinds of knowledge that ought to be the outcome of high school education: natural sciences, foreign languages, history, and so on.

The writer is Dean of Admissions and an associate professor of English at the University of Virginia.

Two facts about this conference ought to matter to Americans who care about children. One is that it happened at all. Most educational leaders now agree that schooling is not working as well as it should. Recent months have seen many attacks on education, especially on our high schools, but few efforts in which colleges and schools come together to begin to solve our mutual problems. The other fact is that the participants came to near-unanimity on what we need to do now to recover what we have lost in schooling.

Thoughtful People

Thoughtful people have argued since the time of Socrates about what we ought to teach. Consensus has been rare, especially in the United States. Now, however, it appears that we educators, from all levels of education, from all parts of the country, and from sundry racial and ethnic backgrounds, may be headed toward a general agreement on how we came to need to improve schooling and on what general goals we ought to set.

The participants heard reports on recent research on the differences between effective and ineffective schooling. Many of the findings contradict popular notions about what has gone wrong. Declining student skills in writing and the other basics, for example, seem not to be the fault of classroom teachers, of the federal government, or of social change. Rather, background research and comments made at the conference hit [illegible] other [illegible] reduced levels of academic rigor in core courses; students choosing to take easy courses when harder courses are available; a rush by colleges to fill dormitories and classrooms at any cost, including sacrifice of basic academic standards.

Higher education's contributions drew particular attention. During the past 15 years, colleges have repeatedly lowered entry requirements in English, history, mathematics, the sciences, and foreign languages. Consequently, many students have acquired the false notion that these basic subjects do not matter. This lack of firm commitment to excellence at the top of the system (in many of our colleges and universities) seems to have taken away the motivation that used to keep students in the hard courses. Most school teachers and counselors have continued to urge students to enroll in the tough courses, but the message has frequently not gotten through.

The recommendations for high schools come in two parts. One part is a description of the basic competencies that any student ought to command before entering college, or for that matter before beginning military service or seeking a first job. The list includes reading, writing, speaking and listening, doing mathematics, reasoning, and studying. Each skill carries with it a brief list of things students ought to be able to perform.

Necessary Skills

Under "reasoning," for example, the conference listed these necessary skills: the ability to identify and formulate problems, as well as the ability to propose and evaluate ways to solve them; the ability to recognize and use inductive and deductive reasoning and to recognize fallacies in reasoning; the ability to draw reasonable conclusions from information found in various sources; the ability to comprehend, develop, and use concepts and generalizations; the ability to distinguish between fact and opinion.

Similar listings were developed for all of the basic competencies. To my way of thinking, these listings of what students ought to be able to do as a result of being educated in high school matter because the skills define what a citizen ought to be able to do. Reasoning of the kind described, for example, seems to me to be essential if we are to vote wisely, to form our own judgments about the world around us, and to think straight. I doubt most students now master it.

Complex List

The kinds of knowledge (as opposed to skills) suggested for students who are to do well in college and meet America's need for leadership in the next generation make a longer, more complex list than most students now attempt. English, mathematics, history, foreign language, science, and the visual and performing arts all received treatment. In each instance, the participants believe the average college-bound student ought to attempt more and accomplish more than he or she does now. Good evidence shows that most high school students are taking fewer courses in mathematics, science, foreign language, and other academic core subjects than was customary 15 or 20 years ago, and that most learn less. The conference calls for a return to complexity in learning.

For Virginia, these tentative conclusions suggest both opportunities and challenges. Our public and private schools still offer the solid, advanced academic courses that seem necessary. Our school superintendents and the State Superintendent of Public Instruction have insisted that our sons and daughters must have the opportunity to excel. Many of our college preparatory schools are models of the academic rigor that pays off for students.

From my experience with our schools, I think our classroom teachers and counselors have remained committed to the core courses. At the same time, our colleges have more often than not failed to set proper standards, and many of our students have chosen the easy courses or the non-academic courses rather than the full program of five or six rigorous courses each year.

In my recent experience, high school juniors and seniors have been especially prone to miss out on what schooling ought to be. Often they either choose partial programs of study, leaving out the tough courses that they need to hone their academic skills, or they begin to earn poor grades just when experience and common sense say they should do their best work.

Sharper Definitions

In our schools and colleges, we might well benefit by developing sharper definitions of goals, higher purposes, and a clearer sense of our obligations. Henry Rosovsky, dean of the Harvard faculty, has observed that students do not know what they do not know, and that they cannot make good choices between options they do not understand. I think he is right. Faculties owe students precisely these kinds of information.

In the end, students and their parents choose what benefit they will get from schooling. We need to do more in our schools than just re-examine and improve their curricula. It may be time, for example, to look into new ways of paying for public education. But the conference report underscores the importance of using resources that already exist in our schools — of taking full advantage of what our tax dollars, tuition payments, and gifts pay for now. Clear writing, intelligent reading, critical thinking: These competencies ought to belong to every citizen. When we choose to demand less of ourselves in school than we should — when we choose to be less able at the end of schooling than we can be — we choose to be less than free and independent. Sound education and effective self-government go hand-in-hand. Neither comes easily.

Mr. **Weiss**. Thank you, Dr. Casteen.

Mr. Sjogren?

STATEMENT OF CLIFF SJOGREN, DIRECTOR OF ADMISSIONS, UNIVERSITY OF MICHIGAN

Mr. **Sjogren**. Mr. Weiss, my name is Cliff Sjogren, director of admissions at the University of Michigan, and president of the American Association of Collegiate Registrars and Admissions Officers.

I appreciate this opportunity to testify on H.R. 1662.

My comments will supplement rather than summarize the more generalized statement that was submitted earlier, and I will speak from the perspective as an undergraduate admissions officer.

The University of Michigan is a large, public supported university that practices a highly selective admissions policy. We are selective because the number of applications from qualified students far exceeds the places available.

The University of Michigan is also characterized by a great diversity in terms of the racial, economic, ethnic, age, and geographical composition of the student population.

AACRAO, founded in 1910, is an important higher educational association with over 7,000 members representing over 2,000 colleges and universities. AACRAO is a member of the ACE secretariat.

I would like to mention that this statement has been prepared after I have thoroughly examined the previous testimony since 1979, both pro and con, and I might add I have learned a great deal from that examination, and my comments today are going to be directed toward those critical and controversial issues that I saw as important in that earlier testimony.

My colleagues at the university and in AACRAO share with the sponsors of H.R. 1662 their wishes to insure equal access for all Americans to educational opportunities of high quality. As we pursue that lofty goal, we are continually reminded of the fragile nature of standard measurements of educational achievement, but we will challenge any effort to remove those evaluative devices.

I will, therefore, speak in opposition to H.R. 1662. Some provisions of that proposed legislation if enacted would severely restrict our ability to carry out our responsibilities to conduct and improve educational access for the citizens of this country.

Specifically, I refer to sections 3(A) 3 and 4 which reveal a serious misunderstanding on how tests are actually used or what admissions officers expect from them and section 5(A) (1) and (2) test disclosure provisions, which will likely reduce test flexibility in terms of the numbers and locations of test sites would also increase costs and add little, I feel, if any, benefits to consumers beyond those now provided by the recently disclosed plan by ETS to disclose test answers.

At first glance an admissions officer might assume that H.R. 1662 is an antitesting effort rather than a consumer interest proposal. Indeed, the catch phrase associated with this effort is "truth in testing," which implies that existing tests are untruthful.

I shall assume that H.R. 1662 is not an antitesting bill and demonstrate, I hope, that consumer interests will best be served by properly used and educationally reliable standardized tests.

As an admissions officer, I hope we can preserve the educational features of testing, however, because I feel a few of the provisions in the proposal are counter to sound educational principles, my testimony will in part be a defense of currently practiced responsible college admissions testing.

A key provision of H.R. 1662 is test item and answer disclosure. If the committee feels that such disclosures serve consumer interests, the recent actions taken by the College Board and ETS in its announced plans for the 1981-82 administration of the Scholastic Aptitude Tests, should satisfy that provision of the proposal.

The College Board reports that students taking the SAT's on five testing dates will have the opportunity to obtain questions, the answers, and their responses on test items. The dates selected will affect approximately 75 percent of all SAT takers in 1981-82. It is a responsible initiative that will allow the ETS to maintain test integrity and important data for test validity. There is little evidence, however, that test item and answer disclosure will "demonstrate" the process.

The College Board found that in New York where similar legislation was passed only 5 percent of the test takers requested the service during the first three administrations. Further, the average median family income of all test takers was $24,500, while that reported by requesters was $32,300.

Orientals and whites requested disclosure at a substantially higher rate than blacks and Puerto Ricans. Of all students who took the SAT's during those three administrations only 2.2 percent of the Puerto Ricans and only 1.5 percent of the blacks requested test item, answer, and their responses be made available to them.

The best predictor of academic behavior is previous academic behavior. Nearly all colleges and universities that employ selective admissions base decisions on a complex and interrelated set of factors that combine to provide evidence of academic potential and educational readiness.

The secondary school record, which includes grades, class rank, selection of courses, and the educational standard of the high school, provide the most significant influence on admission decisions. The entrance test, in varying degrees, will also be an important factor in many cases. Numerous studies have proven that standardized tests when properly used significantly improve the capability of predicting academic success in college.

The importance of the test score diminishes, however, as consistency and quality, both at high and low levels, are revealed in the high school record. It frequently occurs that a student with borderline grades, or one from an unaccredited, rural, or innercity school will demonstrate very high promise by scoring well on a test.

That student would probably be overlooked if test scores were not a part of the process. Admissions officers generally consider low test scores as ambiguous and high scores, on reliable tests, as relatively free of ambiguities. For minority students, for poor students, and rural students coming out of very small high schools, test scores, it has been our experience, will tend to work for those students rather than against those students.

Reliable test data helps hold back the negative influences of grade inflation. It would be impossible for the University of Michigan, for example, to assess the quality of the more than 1,000 high schools from which applications are received. Standardized measurements are needed to maintain our efforts to treat fairly all candidates for admission and to not rely solely on subjective teacher evaluations.

Interestingly, if we were to remove tests from the admissions decision, it would be likely that students coming out of the preparatory schools, private, prestigious schools, would have a clear advantage over students coming out of schools where counselors were less sophisticated, lower paid, and did not have the ability to articulate thoughts on evaluations and the types of commentary that we typically get out of the more suburban schools.

The admissions process at Michigan, therefore, depends on and is strengthened by a relatively low cost, frequently administered examination plan that will allow students to demonstrate their aptitudes and achievements without coming to Ann Arbor for a locally constructed and administered entrance test. Standardized examinations were originally designed with that convenience for the student in mind.

Also, test information, when properly displayed in an institution's preapplication literature, will allow and encourage students to "select a college." One might argue that the ultimate in consumerism is to require institutions to display a profile on which are described important institutional characteristics, thereby allowing self-selection to become the single most significant factor in the process.

Finally, a few comments on some things that admissions officers agree tests cannot do. Generally, college entrance tests cannot help predict success in a career. Tests are of little value in predicting academic achievement beyond one year of college, but for freshmen, particularly, that one year is critical.

Tests are almost useless when used as a sole criterion for admission. A recent AACRAO/College Board survey revealed that less than 2 percent of U.S. higher educational institutions consider tests as the single most important criterion in the selection process. Probably none consider tests as the only criterion.

Tests should not, and generally are not, interpreted as specific scores but rather as ranges within which falls the margin of error. Thus, it is unlikely that the occasional incorrect answer that surfaces will influence an admissions decision.

Tests administered by companies that do not provide students, counselors, and admissions personnel with complete descriptions as generally described in H.R. 1662, section 3, with some exceptions in provisions section 3A (3) and (4), should be removed from the market and the companies responsible should be punished.

Failure to obtain admission to a college of first choice because of unacceptable academic qualifications will almost always be disappointing to the applicant and sometimes traumatic. Dismissal from a college because of insufficient ability or preparation will almost always be traumatic and sometimes tragic.

Properly developed and administered tests when fully understood by the student and admissions officer will greatly increase the probability that the college choice will be a wise one. There have been abuses in the past, and there will undoubtedly be some, although fewer, I suspect, in the future.

The responsible testing agencies, however, have responded quickly and positively to the issue of test misuse. Admissions personnel have been increasingly knowledgeable on what tests can and cannot do and are adjusting their practices accordingly. And the public attention that has focused on this critical issue in recent months will likely flush out further abuse.

Past experience has shown that when well-intentioned laws are submitted to the regulatory agencies for the preparation of the implementation plan, the intent of Congress is frequently misunderstood or ignored.

Traditionally, education in the United States has been relatively free of unnecessary political influence. That is a tradition which must be preserved as it has served this country well.

We would do well to remember that at the turn of the century liberal and progressive thinkers demanded that a national testing plan be devised to encourage college access on the basis of demonstrated ability and not as was practiced at that time by chance of birth or attendance at the proper prep schools.

Let us not return to the 19th century.

Thank you.

[The prepared statement of Cliff Sjogren follows:]

I submit this statement for the record first as the Director of Admissions at the University of Michigan, a large institution of higher learning that consists of eight undergraduate schools and colleges and ten graduate and professional schools, each of which employs selective and competitive admissions.

I also testify on behalf of the American Association of Collegiate Registrars and Admissions Officers (AACRAO), a national, non-profit educational association which is composed of approximately 2,000 accredited colleges and universities that are represented by over 7,100 individual members from the field of recruitment, admissions, registration and records, student financial aid, and institutional research at the undergraduate, graduate, and professional levels. I currently serve as president of the association.

My colleagues and I share with the sponsors of HR 1662 their basic philosophy of providing equal and the best possible higher education opportunities for all of the youth of our nation--and I might add we support these opportunities for all adults as well. We do not share with the sponsors and the proponents of this proposed legislation the thesis that tests are the sole or primary criterion for admissions to higher education.

The Executive Committee of AACRAO has reviewed carefully the controversy surrounding the interpretation of admissions tests as discussed in the Nader/Nairn report, particularly charges that admissions test scores are grossly misused as inhibitors in undergraduate admissions decisions. AACRAO's professional journal, College and University, is replete with numerous research articles emphasizing a balance of criteria that should be reviewed in the admissions process, e.g., high school grades, counselors recommendations, leadership potential, indications of latent academic achievement etc., as well as admissions test scores. Also, these journal articles particularly stress the proper interpretation of admissions tests in determining the potential of success for minority and disadvantaged students. The Association's

RECENT COOPERATIVE STUDY ON ADMISSIONS POLICIES AND PRACTICES IN THE 1980'S CLEARLY INDICATES THAT ADMISSIONS TEST SCORES ARE USED IN ISOLATION ONLY IN A RELATIVELY SMALL NUMBER OF INSTITUTIONS.

THE AACRAO EXECUTIVE COMMITTEE REVIEWED A PAPER PREPARED BY A SUB-COMMITTEE OF ITS PROFESSIONAL COMMITTEE ON UNDERGRADUATE ADMISSIONS. AACRAO ENDORSES THIS STATEMENT AS INDICATIVE OF THE ASSOCIATION'S PERSPECTIVE ON THE CONTROVERSY ABOUT ADMISSIONS TESTING AND THE USE OF TESTS IN THE ADMISSIONS PROCESS.

A CURRENT PERSPECTIVE ON STANDARDIZED TESTING IN UNDERGRADUATE ADMISSIONS

AN AACRAO POSITION PAPER PREPARED BY A SUBCOMMITTEE OF THE
COMMITTEE ON UNDERGRADUATE ADMISSIONS

AS AMERICAN HIGHER EDUCATION HAS MATURED, IT HAS BECOME MARKED BY INCREASINGLY DIVERSE MODELS OF INSTITUTIONAL PURPOSE, OF GOVERNANCE, AND OF ACCESS BY STUDENTS. PERHAPS THE EVIDENCE OF THESE TRENDS MOST VISIBLE TODAY IS A GENERATION OF STUDENTS WHO ARE MORE DISSIMILAR TO EACH OTHER SOCIALLY, ECONOMICALLY, AND INTELLECTUALLY THAN ANY WHICH HAS PRECEDED. THIS DIVERSITY REFLECTS A BROAD SPECTRUM OF ADMISSIONS PRACTICES WHICH RANGE FROM SKIMMING ONLY A SMALL FRACTION FROM A HIGHLY QUALIFIED APPLICANT POOL TO ADMITTING ANYONE PAST A GIVEN AGE OR WITH ANY KIND OF EDUCATIONAL HISTORY.

FEW OF AMERICA'S NEARLY THREE THOUSAND COLLEGES HAVE ADMISSIONS POLICIES WHICH ARE EXACTLY ALIKE. BY AND LARGE, THESE DIFFERENCES ARE NOT ARBITRARY, SIMPLISTIC, OR CAPRICIOUS, BUT INDEED REFLECT REASONED CONSTRAINTS BASED ON RESOURCES AVAILABLE, PHILOSOPHIES OF INSTRUCTION, QUALITY AND MISSION OF THE RESPECTIVE FACULTY, AND COMMITMENTS TO VARIOUS CONSTITUENCIES. JUST AS THESE CONSIDERATIONS ARE DIVERSE, SO IS THE DESIRED MIX OF STUDENT ATTRIBUTES WHICH THE MANY COLLEGES PURSUE, AND UPON WHICH ADMISSIONS DECISIONS ARE JUSTIFIED.

TYPICALLY, SELECTIVE ADMISSION IS BASED ON A COMPLEX AND INTERRELATED SET OF FACTORS WHICH COMBINE TO PROVIDE CONVINCING EVIDENCE OF ACADEMIC PROMISE AND EDUCATIONAL READINESS, ALTHOUGH COLLEGES WITH SPECIAL PURPOSES MAY ALSO VALUE SUCH OTHER QUALITIES AS RELIGIOUS CONVICTION, INTERESTS, AND CHARACTER TRAITS. THE BEST PREDICTOR OF

ACADEMIC BEHAVIOR LONG HAS BEEN ACKNOWLEDGED AS PREVIOUS ACADEMIC BEHAVIOR. USUALLY THIS CONSISTS OF PERFORMANCE ABSTRACTED FROM THE HIGH SCHOOL RECORD, NOTABLY THE NATURE OF COURSES TAKEN, GRADES RECEIVED, AND RANK IN THE GRADUATING CLASS.

NONETHELESS, AS STRONG A PREDICTOR OF SUCCESS AS THE SECONDARY RECORD IS, IT IS FAR FROM PERFECT. QUALITATIVE DIFFERENCES AMONG HIGH SCHOOLS ACROSS THE NATION, YEAR TO YEAR FLUCTUATIONS AMONG GRADUATING CLASSES WITHIN SCHOOLS, VARIATIONS IN GRADING PRACTICES AMONG TEACHERS AND SCHOOLS, AND MARKED INCONSISTENCIES WITHIN A STUDENT'S OWN RECORD ALL HAVE BORNE OUT THE NEED FOR A SECOND MEASURING DEVICE WHICH COMPARES THE ACADEMIC ATTRIBUTES OF INDIVIDUAL CANDIDATES WITH THOSE OF A NATIONAL COLLEGE-BOUND POPULATION SPANNING MANY YEARS.

ADMINISTRATORS CHARGED WITH MAKING RESPONSIBLE ADMISSIONS DECISIONS LONG HAVE WELCOMED THE EQUALIZING INFLUENCE THAT TESTS WITH WELL ESTABLISHED PROPERTIES OF STABILITY AND PREDICTIVENESS ADD TO THE USE OF THE SECONDARY RECORD ALONE. WHILE THE IMPROVEMENT THAT STANDARDIZED TESTS LEND TO THE PREDICTION OF FIRST YEAR PERFORMANCE IS NOT ALWAYS LARGE, IT CONSISTENTLY AND SIGNIFICANTLY ENHANCES THE PROBABILITY THAT AT LEAST THE EARLY PHASE OF THE APPLICANT'S ACADEMIC CAREER WILL BE A SUCCESSFUL ONE.

IN PREDICTING PERFORMANCE, COLLEGES MAY COMBINE SUMMED OR AVERAGED SCORES FROM WHATEVER BATTERY OF TESTS IS AVAILABLE WITH PREVIOUS GRADES, OR, WHEN PREDICTION IS MORE SOPHISTICATED OR SPECIALIZED, GIVE ADDITIONAL WEIGHTING TO CERTAIN SCORES OR SUBSCORES. SIMILARLY, A NUMBER OF SUCH SCORES MAY BE VIEWED AS A PROFILE OF INDIVIDUAL ABILITIES, WHERE RELATIVE STRENGTHS MAY BE EVEN MORE REVEALING AND HELPFUL TO THE STUDENT IN ADVISING, PLACEMENT, AND GRANTING CREDIT - EQUALLY IMPORTANT DIMENSIONS OF THE ADMISSIONS FUNCTIONS.

STANDARDIZED TESTS ALSO ARE INSTRUMENTAL IN THE STUDENT'S CHOICE OF A COLLEGE - A DYNAMIC WHICH IS FAR MORE POTENT IN NATIONAL ENROLLMENT PATTERNS THAN IS THAT OF COLLEGES CHOOSING STUDENTS. TYPICALLY, THE PROSPECTIVE STUDENT CONFRONTS A NUMBER OF POTENTIAL INSTITUTIONS, NARROWS THESE TO THE ONES TO WHICH APPLICATION IS MADE, THEN CHOOSES AMONG THOSE WHICH EXTEND THE OFFER OF ADMISSION. THE INTERPRETATIONS WHICH STUDENTS AND THEIR ADVISORS ATTACH TO INDIVIDUAL SCORES WITHIN THE CONTEXTS OF PREDICTED SUCCESS IN THE PROSPECTIVE COLLEGES ARE VITAL ELEMENTS AT EACH STAGE OF THE DECISION

PROCESS. MOREOVER, THE EFFECTIVENESS WITH WHICH PROSPECTIVE STUDENTS CAN COMMUNICATE WITH A NUMBER OF INSTITUTIONS UNDER CONSIDERATION IS GREATLY FACILITATED BY THE NATIONAL TESTING PROGRAMS, WHO DISSEMINATE AT VERY REASONABLE COST TO THE STUDENT NOT ONLY TEST DATA WHICH IS NATIONALLY COMPARABLE AND READILY INTERPRETED, BUT A BODY OF PERSONAL AND ACADEMIC INFORMATION ALSO COLLECTED DURING THE TESTING PROCESS.

WERE THESE SERVICES NOT AVAILABLE, COLLEGES AND STUDENTS ALIKE WOULD FACE PROHIBITIVE COSTS OF COMMUNICATING INDIVIDUALLY WITH A LARGE NUMBER OF INSTITUTIONS, SUBMITTING SETS OF INFORMATION UNIQUE TO EACH, AND SITTING FOR A NUMBER OF LOCAL EXAMINATIONS WHICH CAN EVALUATE APPLICANTS ONLY WITHIN A LIMITED CONTEXT. OUT OF COST AND INCONVENIENCE ALONE, OPTIONS FOR STUDENTS WOULD SEVERELY BE CURTAILED - AN INTOLERABLE SETBACK TO FREEDOM OF CHOICE AND DEMOCRATIZATION OF ACCESS. AND WERE THERE NOT AVAILABLE THIS MEANS OF DISTRIBUTING TO A NUMBER OF COLLEGES SUCH INFORMATION RELEVANT TO PREDICTED ACADEMIC PERFORMANCE AND THE DESIRED CAMPUS MIX, THEY, TOO, WOULD FIND THEIR FREEDOM OF CHOICE TO SEARCH FOR, TO ATTRACT, AND TO SERVE STUDENTS SIMILARLY CONSTRAINED. PERHAPS THE MOST CONVINCING EVIDENCE OF THE VALUE OF THIS SERVICE HAS BEEN THE OPPORTUNITIES FOR MINORITY AND OTHER DISADVANTAGED STUDENTS WHICH HAVE INCREASED COINCIDENTLY WITH THEIR PARTICIPATION IN THE ADMISSIONS TESTING PROGRAMS.

SO THAT THE USE AND UNDERSTANDING OF ADMISSIONS TEST INFORMATION MIGHT CONTINUALLY BE ENHANCED, THE TEST PUBLISHERS ADVOCATE THAT COLLEGES SYSTEMATICALLY EVALUATE THE PREDICTIVE VALIDITY OF THE FACTORS EACH USES IN THE SELECTION OF STUDENTS. CONSEQUENTLY, THE RELATIVE WEIGHTING OF GRADES, SCORES, AND OTHER MEASURABLE QUALITIES CAN BE ADJUSTED EMPIRICALLY TO CHANGES IN THE SOCIAL AND EDUCATIONAL MILIEU, AS WELL AS TO CHANGES IN THE COLLEGES' OWN AIMS. SO LONG AS OUR HIGHER INSTITUTIONS CAN DEMONSTRATE TO THEIR OWN SATISFACTION THAT WHATEVER DEVICES THEY EMPLOY BRING ABOUT THE MOST EFFECTIVE MATCH AMONG STUDENT ABILITIES, ACADEMIC SUCCESS, AND THE EDUCATIONAL RESOURCES AVAILABLE, AACRAO AFFIRMS THAT THE USE OF STANDARDIZED TESTS IS A LEGITIMATE AND VALUABLE COMPONENT IN THE EVALUATION AND SELECTION OF UNDERGRADUATE STUDENTS.

IT IS NOT AN UNCOMMON TENDENCY FOR THE LAYMAN TO MISTRUST A TECHNOLOGY WHICH HE OR SHE DOES NOT FULLY UNDERSTAND, AND TO CASTIGATE IT IN THE EVENT OF ACCIDENTAL OR INTENDED MISUSE. BUT A PROPERLY REFINED TECHNOLOGY IS BLAMELESS: IT IS PEOPLE WHO USE TESTS, AND WHOSE PROFESSIONAL ACUMEN MUST BE ADEQUATE TO THE STRENGTHS AND LIMITATIONS OF THE TOOLS THEY EMPLOY. AND IT IS THE EDUCATED JUDGMENTS, HUMANE AND SCIENTIFIC, ON THE PART OF THOSE WHO MATCH STUDENTS WITH INSTITUTIONS WHICH WILL CONTINUE TO IMPROVE EDUCATIONAL OPPORTUNITIES FOR ALL WHO SEEK THEM.

Mr. WEISS. Thank you, Dr. Sjogren.

Dr. Robinson?

STATEMENT OF SHARON ROBINSON, DIRECTOR, INSTRUCTION AND PROFESSIONAL DEVELOPMENT, NATIONAL EDUCATION ASSOCIATION

Ms. ROBINSON. I am Sharon Robinson, director of instruction and professional development for the National Education Association.

I am pleased to testify today on behalf of the NEA and in support of H.R. 1662, the Educational Testing Act of 1981.

Before I get into the remarks that I would like to share with the panel this morning, let me take an opportunity to perhaps clear up a common misunderstanding.

I would like to share with you the NEA resolution on testing.

The National Education Association recognizes that testing of students preschool through job entry may be appropriate for such purposes as identifying learning needs, recommending instructional activities, and describing opportunity progress.

The association opposes the use of tests that deny students full access to equal educational opportunities or that are used to evaluate teachers.

The association believes that standardized tests should not be administered when they are potentially damaging to students' self-concept, bias, used as the only criterion for student placement, invalid, unreliable, or out-of-date, used as a basis for the allocation of Federal, State, or local funds, used by testing companies or publishers to promote their own financial interest at the expense of sound educational uses, used to compare individual schools, used in an exploitive manner by the media, used as the sole criterion for graduation or promotion, and appropriate for the use intended, and last, used as a criterion for the development of a State system of classification of schools and/or school systems.

Clearly, the NEA has recognized that tests have their place. We do not dispute that testing has value, and that through testing we are aided in making some very difficult discriminating choices by students and by educators.

But too often test scores result in an assumptive pronouncement about students, about schools, and about educators. Such a pronouncement is based solely on those scores as it often is likely to be invalid and incorrect.

Clearly, the measure of accountability, some measure of accountability for the testing industry and for the uses of tests is long overdue. NEA research confirms that test usage [standardized test usage] is pervasive throughout this country, and a recent survey in 1980 of various school systems, the sample reported 81 percent of the school districts are using some form of group-administered standardized intelligence tests.

Usage of some group-administered standardized achievement test was reported by 99 percent of the sample. Ninety-five of those districts reported using group-administered, standardized, reading-readiness tests; 91 percent of those districts use group-administered aptitude tests, and 89 percent used some form of group-administered standardized interest tests.

I will supply a written copy of this report to the reporter for the record.

Mr. WEISS. Without objection, that will be entered.

[The information referred to follows:]

RESEARCH MEMO

National Education Association
1201 Sixteenth Street, N.W. • Washington, D.C. 20036

Survey of School Programs and Practices 1980

NEA Research Memo *September 1980*

SURVEY OF SCHOOL PROGRAMS AND PRACTICES—1980

This 1980 survey of school systems regarding some of their programs and practices was sent to 805 school systems. The school systems were selected to be representative of the approximately 11,600 public school systems in the United States with 300 or more students. Exactly 508 usable questionnaires were returned for a 63.1 percent response. The standard error for population estimates is 5 percent or less and for estimating population differences is 10 percent or more.

Each topical area includes highlights of the responses plus a more detailed analysis of each question by region and by size of school system. The topical areas are as follows:

Numbers of students and teachers
School organization patterns
Curriculum
Instruction in special areas
Student evaluation
Student discipline
Financing
Censorship
Handicapped students
Professional development
Teacher evaluation
Discrimination
Specialists

The states comprising each region are as follows:

Northeast: Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, Pennsylvania, Rhode Island, and Vermont.

Southeast: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia.

Middle: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin.

West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, and Wyoming.

School systems by stratum are *Small:* 300-2,999 students; *Medium:* 3,000 to 24,999 students; *Large:* 25,000 or more students.

Numbers of Students/Teachers

- School systems in the Southeastern and Western regions are generally larger than school systems in the Northeastern and Middle regions—in fact, more than twice as large on the average.

- Almost two-thirds (66 percent) reported that student enrollments have decreased over the past three years (i.e., since 1976-77). More school systems in the Northeastern and Middle regions reported a decrease than did school systems in the Southeastern and Western regions.

- Decreases in the number of teachers (reported by 41 percent) were accomplished primarily by normal attrition (81 percent) rather than by reduction in force (10 percent).

- About half felt that the number of students (52 percent) and number of teachers (48 percent) should decrease within the next three years (i.e., by 1982-83). Only about one-sixth (17 percent) felt the number of students and teachers should increase. Northeastern and large school systems were more likely to think that the number of both students and teachers would decrease in the next three years.

NUMBERS OF STUDENTS/TEACHERS

Question	Total	Region Northeast	 Southeast	 Middle	 West	Stratum 300-2,999	 3,000-24,999	 25,000-or more
What was the total number of teachers in your school system as of October 1979? (Percent responding.)								
Elementary (K-6):								
Less than 30	16	12	6	20	15	48	0	0
30 to 49	9	12	4	9	8	27	0	0
50 to 99	17	27	14	16	17	23	23	0
100 to 199	17	29	22	14	9	2	42	0
200 to 499	13	9	14	13	16	0	29	6
500 to 999	12	5	8	13	17	0	2	40
1,000 or more	16	5	34	14	19	0	3	55
Mean	696	228	1,531	439	1,114	44	283	2,102
Median	129	98	247	121	228	31	149	1,039
Secondary (7-12):								
Less than 30	15	8	5	20	16	46	0	0
30 to 49	10	15	5	10	11	28	3	0
50 to 99	15	17	13	16	13	23	18	0
100 to 199	18	33	24	15	10	1	44	1
200 to 499	13	14	11	12	15	1	27	6
500 to 999	15	6	15	15	21	0	5	48
1,000 or more	14	7	27	12	16	0	3	45
Mean	609	281	1,187	414	879	41	289	1,751
Median	140	116	216	128	205	34	150	941

NUMBERS OF STUDENTS/TEACHERS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Please give the October 1979 ENROLLMENTS requested below Do not include enrollment in adult education and junior college.								
a. Nursery and prekindergarten								
Percent responding	27	33	24	24	32	10	30	44
Mean	398	105	523	280	743	33	69	832
Median	86	44	192	82	90	15	41	303
b. Kindergarten and prefirst grade								
Percent responding	90	90	96	93	86	85	90	97
Mean	1,433	540	2,485	970	2,303	81	442	4,299
Median	312	215	623	264	616	62	331	2,689
c. Elementary (1-6)								
Percent responding	94	93	96	97	88	89	95	97
Mean	8,646	3,568	15,709	6,690	11,908	551	3,073	26,220
Median	2,100	1,600	4,472	1,734	4,231	437	2,395	17,553
d. Secondary (7-12)								
Percent responding	95	93	98	96	95	89	98	99
Mean	8,775	4,028	15,970	6,571	11,560	565	3,356	26,247
Median	2,241	1,800	4,082	1,750	4,429	486	2,354	17,909
To what extent have student enrollment and number of teachers changed in your school system over the *past three years* (since the 1976-77 school year)? (Percent responding.)								
a. Student enrollment								
Increased	16	9	12	8	30	20	15	10
Median percent of increase	8	6	8	7	10	10	7	7
Relatively no change	19	17	29	18	15	26	20	11
Decreased	66	74	59	73	55	55	66	78
Median percent of decrease	9	8	10	11	7	9	8	8
b. Number of teachers								
Increased	25	14	36	17	34	25	26	21
Median percent of increase	7	7	7	6	10	8	7	6
Relatively no change	34	31	34	40	34	47	28	29
Decreased	41	56	30	43	32	28	47	50
Median percent of decrease	6	5	10	7	5	6	7	6
Generally, how has any decrease in the number of teachers occurred—by normal attrition or by "reduction in force"? (Percent responding.)								
Numbers have NOT decreased	48	35	60	41	57	63	44	34
All by normal attrition	32	27	24	39	27	22	29	47

NUMBERS OF STUDENTS/TEACHERS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
About three-fourths by normal attrition	10	18	4	10	10	6	13	11
About half by attrition and about half by "reduction-in force"	5	9	8	3	4	4	7	11
About three-fourths by "reduction in force"	3	8	1	4	0	2	6	2
All by "reduction in force"	2	4	3	2	2	3	2	2
How does your school system estimate that student enrollment and numbers of teachers will change in the next three years (by the 1982-83 school year)? (Percent responding.)								
a. Student enrollment								
Should increase	17	8	14	12	31	23	18	10
Should remain relatively constant	31	28	35	37	23	41	28	22
Should decrease	52	65	51	52	46	36	54	68
b. Number of teachers								
Should increase	17	7	18	10	29	20	18	11
Should remain relatively constant	36	30	42	40	29	48	29	31
Should decrease	48	63	41	50	43	32	54	58

School Organization Patterns

- One-fourth (25 percent) of the school systems reported a 6-3-3 school organization for grades 1 through 12. 5-3-4 and 6-2-4 organizations were reported by about one-sixth (15 and 19 percent, respectively).
- The 5-3-4 and 6-3-3 organizations were most common in the Northeastern region, and the 6-3-3 organization most common in the Southeastern region, and the 6-2-4 or 6-3-3 organization most common in the Middle and Western regions.
- The larger school systems were more likely to have a 6-3-3 organization, with a significant number having a 6-2-4 organization, whereas the smaller school systems were more likely to have a 6-6 organization, with a significant number having a 6-2-4 or 8-4 organization.

SCHOOL ORGANIZATION PATTERNS

		Region				Stratum		
Question	Total	Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
What type of SCHOOL ORGANIZATION by grade (not including kindergarten) is the *most general* practice in your school system?								
3-3-3-3	1	3	1	1	1	1	1	2
3-3-6	1	1	1	0	0	1	1	0
4-4-4	6	8	4	7	4	11	6	2
5-3-4	15	23	18	16	15	8	16	23
6-2-4	19	13	9	20	21	16	22	17
6-3-3	25	18	33	23	26	8	30	38
6-6	9	11	10	12	3	22	4	0
7-5	2	1	5	1	1	2	1	3
8-4	9	4	6	13	8	16	5	5
Other	14	20	12	9	20	16	15	11

The Curriculum

- A majority of the school systems said their school system has felt pressure to "enforce stricter discipline" (61 percent). That pressure was less likely to be felt in smaller school systems. A majority of the large school systems were also likely to feel pressure to "increase the time spent on the basic curriculum" (73 percent), "require higher student achievement standards" (74 percent), "require students to pass a test before being promoted or graduated" (59 percent), and to "enforce stricter attendance requirements" (62 percent).

- At the elementary level, reading, American history, general science, mathematics/arithmetic, physical education, and spelling are required by practically every school system. Citizenship, geography, and state history are required by more than 85 percent. Music is required by 81 percent, world history by 78 percent, and fine arts by 70 percent. A foreign language is not even offered by 70 percent of the school systems.

- At the junior high level, almost every school system requires English, science, mathematics, and physical education. American history is required by 92 percent, and world history and state history are required by more than two-thirds (67 percent and 78 percent, respectively). Fine arts, instrumental and vocal music, home economics, and industrial arts are offered by most school systems, but generally as an elective. About half offer typing (56 percent) and vocational education (48 percent) generally as an elective. As an elective, Spanish is offered by 61 percent, French by 54 percent, and German by 28 percent. Very few junior high schools offer Russian or Chinese.

- At the high school level, English, American history, and mathematics are the only courses required by most of the school systems. A significant number require world history (52 percent), biology (44 percent), and/or another science (60 percent), while most of the other school systems offer them as an elective. A wide range of other courses is offered as an elective by most school systems with the exception of a state history offered by 68 percent, German by 63 percent, Russian by 22 percent, Chinese by 8 percent and agriculture by 61 percent.

THE CURRICULUM

		Region				Stratum		
Question	Total	Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
There has been a much-publicized demand for schools to get "back to the basics." To what extent has your school system felt pressure to do each of the following? (Percent responding "much" or "moderate" pressure.)								
a. Increase the time spent on the basic curricula	49	49	58	43	54	34	37	73
b. Increase the amount of student homework	28	30	28	25	35	15	25	49

THE CURRICULUM—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
c. Structure the schools in the "traditional" manner	28	32	30	22	30	19	28	38
d. Enforce strict discipline	61	61	63	62	61	52	59	74
e. Require higher student achievement standards	57	54	65	56	57	39	58	77
f. Require students to pass a test before being promoted or graduated	35	27	60	26	38	15	36	59
g. Enforce strict attendance requirements	46	39	53	42	50	36	45	62
Which of the following COURSES does your school system consider as basic subjects; that is, they are required of all *regular* students for graduation from the public schools? (Percent responding.)								
ELEMENTARY:								
a. Reading								
Required	100	100	100	99	100	100	100	99
Elective	0	0	0	1	0	0	0	1
Not offered	0	0	0	0	0	0	0	0
b. Citizenship								
Required	88	95	83	87	87	87	89	89
Elective	4	0	7	2	5	4	4	2
Not offered	8	5	10	11	8	9	7	8
c. Geography								
Required	91	95	88	95	86	96	86	94
Elective	5	1	8	3	7	2	8	3
Not offered	4	4	4	2	7	2	5	3
d. American history								
Required	95	100	92	96	93	94	95	97
Elective	1	0	3	1	1	1	2	1
Not offered	4	0	5	3	6	6	4	2
e. World history								
Required	78	86	79	77	79	77	80	76
Elective	7	7	7	8	1	7	6	8
Not offered	15	7	14	16	20	16	14	16
f. State history								
Required	87	91	91	89	82	89	85	87
Elective	4	0	4	3	2	3	3	5
Not offered	9	9	5	9	16	8	12	8

THE CURRICULUM—(CONTINUED)

		Region				Stratum		
Question	Total	Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
g. General science								
Required	98	98	99	98	98	96	98	98
Elective	1	3	0	1	0	1	1	2
Not offered	1	0	1	1	2	3	1	1
h. Mathematics/arithmetic								
Required	100	100	100	100	100	100	100	100
Elective	0	0	0	0	0	0	0	0
Not offered	0	0	0	0	0	0	0	0
i. Fine arts								
Required	70	82	64	74	63	70	68	73
Elective	21	9	23	20	27	18	22	24
Not offered	9	9	12	7	10	12	10	2
j. Music								
Required	81	92	71	86	78	81	80	82
Elective	17	8	27	12	19	18	17	18
Not offered	2	0	3	2	2	2	3	0
k. Physical education								
Required	98	99	95	98	100	96	98	99
Elective	1	0	4	1	0	1	1	1
Not offered	1	1	1	2	0	3	1	0
l. A foreign language								
Required	5	5	5	4	6	5	4	7
Elective	25	14	39	17	35	20	19	42
Not offered	70	82	57	79	59	75	77	51
m. Spelling								
Required	97	96	99	98	99	98	96	98
Elective	2	1	1	2	0	1	3	1
Not offered	1	3	0	1	1	1	1	1
JUNIOR HIGH:								
a. English								
Required	100	100	100	100	100	100	100	100
Elective	0	0	0	0	0	0	0	0
Not offered	0	0	0	0	0	0	0	0
b. American history								
Required	92	95	88	91	94	92	90	94
Elective	4	1	3	4	2	3	4	2
Not offered	5	4	10	5	4	5	6	4
c. World history								
Required	67	79	64	66	65	71	72	57
Elective	13	9	20	13	10	10	11	20
Not offered	19	12	16	21	24	19	17	24
d. State history								
Required	78	82	73	77	82	83	75	74
Elective	7	4	12	7	4	4	9	8
Not offered	15	14	15	16	15	12	16	18

THE CURRICULUM—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
e. A science								
Required	98	97	99	99	98	99	98	97
Elective	2	3	1	1	1	1	2	3
Not offered	0	0	0	0	1	0	1	0
f. Mathematics								
Required	99	98	100	100	100	100	99	100
Elective	1	1	0	0	0	0	1	0
Not offered	0	0	0	0	0	0	0	0
g. Physical education								
Required	98	99	96	98	99	97	97	98
Elective	2	0	4	3	1	3	3	2
Not offered	0	1	0	0	0	1	0	0
h. Fine arts								
Required	50	81	32	54	40	55	54	38
Elective	44	14	61	39	58	35	44	60
Not offered	6	5	7	7	2	10	2	2
i. Instrumental music								
Required	7	15	4	6	6	10	8	4
Elective	90	83	88	91	93	86	89	96
Not offered	3	3	8	3	1	5	3	0
j. Vocal music								
Required	26	46	16	30	18	27	26	23
Elective	70	53	75	66	79	63	72	77
Not offered	4	1	9	5	4	10	3	0
Foreign language:								
k. Spanish								
Required	2	3	0	1	2	1	2	2
Elective	59	64	71	44	63	28	63	89
Not offered	39	33	29	54	33	71	35	9
l. French								
Required	2	4	0	2	1	2	3	1
Elective	52	67	69	43	48	18	55	84
Not offered	47	29	31	55	51	80	43	15
m. German								
Required	1	2	0	1	0	1	1	1
Elective	27	26	30	23	31	10	25	52
Not offered	72	72	70	76	69	90	74	47
n. Russian								
Required	0	0	0	1	0	0	1	0
Elective	5	7	10	2	4	0	2	17
Not offered	95	94	90	97	96	100	97	84
o. Chinese								
Required	0	0	0	1	0	0	1	0
Elective	3	2	3	1	3	0	1	10
Not offered	97	98	97	98	97	100	99	90

THE CURRICULUM—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
p. Home economics								
Required	34	64	16	36	28	37	37	25
Elective	59	32	77	56	66	47	57	75
Not offered	8	4	7	8	6	16	6	0
q. Typing								
Required	4	4	0	5	4	6	3	2
Elective	52	38	75	46	52	29	51	81
Not offered	44	58	25	49	45	65	46	18
r. Agriculture								
Required	1	2	0	2	0	4	1	0
Elective	22	11	41	17	21	20	18	29
Not offered	77	88	59	81	79	76	81	71
s. An industrial art								
Required	33	65	15	36	27	34	38	23
Elective	56	30	71	52	66	46	52	76
Not offered	11	5	15	12	7	20	10	1
t. Vocational education								
Required	5	7	4	5	4	6	4	4
Elective	43	26	69	38	44	25	42	66
Not offered	52	67	27	58	53	59	54	30
u. Driver education								
Required	3	2	4	3	1	3	2	3
Elective	18	12	24	18	10	19	17	16
Not offered	80	87	72	79	89	77	81	82
SENIOR HIGH:								
a. English								
Required	100	100	100	100	99	100	100	100
Elective	0	0	0	0	0	0	0	0
Not offered	1	0	0	0	1	0	1	0
b. American history								
Required	98	100	99	98	99	99	98	98
Elective	1	0	1	3	0	1	1	2
Not offered	0	0	0	0	1	0	1	0
c. World history								
Required	52	61	57	41	55	53	59	42
Elective	47	39	42	57	44	45	40	58
Not offered	1	0	1	2	1	2	1	0
d. State history								
Required	35	46	29	32	42	41	33	30
Elective	33	26	34	31	34	24	34	42
Not offered	32	29	37	37	25	35	33	28

THE CURRICULUM—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000- or more
e. Biology								
Required	44	44	55	40	41	50	44	38
Elective	55	56	45	60	58	50	56	62
Not offered	1	0	0	0	1	0	1	0
f. Another science								
Required	60	63	66	48	71	60	59	60
Elective	40	37	34	52	29	39	40	40
Not offered	1	0	0	1	1	1	1	0
g. Mathematics								
Required	91	86	96	89	96	90	91	95
Elective	8	14	4	11	3	10	9	5
Not offered	1	0	0	0	1	0	1	0
h. Fine arts								
Required	14	22	8	11	16	19	14	9
Elective	82	77	82	85	82	74	83	91
Not offered	4	1	10	4	2	7	3	0
i. Performing arts (e.g., drama)								
Required	1	0	0	1	1	1	1	1
Elective	91	91	88	88	96	84	92	99
Not offered	9	9	12	11	3	15	8	0
j. Instrumental music								
Required	1	0	0	1	2	1	1	1
Elective	98	100	96	97	98	97	98	99
Not offered	1	0	4	1	0	2	1	0
k. Vocal music								
Required	1	0	1	2	0	1	1	2
Elective	96	97	91	95	97	91	97	98
Not offered	4	1	8	3	3	8	2	0
Foreign language:								
l. Spanish								
Required	2	0	1	2	1	1	2	2
Elective	87	97	87	80	95	68	95	98
Not offered	11	3	12	18	3	31	4	0
m. French								
Required	1	0	1	2	0	2	1	1
Elective	83	99	86	79	80	53	95	99
Not offered	15	2	13	20	20	47	4	0
n. German								
Required	1	0	2	2	0	2	1	1
Elective	62	79	55	51	70	25	70	91
Not offered	37	21	43	48	30	73	29	8
o. Russian								
Required	0	0	0	0	0	1	0	0
Elective	22	33	30	11	21	3	15	53
Not offered	78	67	70	89	79	97	85	47

THE CURRICULUM—(CONTINUED)

Question	Total	Region: Northeast	Southeast	Middle	West	Stratum: 300-2,999	3,000-24,999	25,000-or more
p. Chinese								
Required	0	0	0	0	0	1	0	0
Elective	8	8	12	3	7	2	3	23
Not offered	92	92	88	98	93	98	97	77
q. Home economics								
Required	4	6	0	4	5	8	3	1
Elective	96	94	100	96	94	92	97	99
Not offered	0	0	0	0	1	1	0	0
r. Typing								
Required	3	3	1	3	3	8	2	1
Elective	97	97	99	97	96	92	98	99
Not offered	0	0	0	0	1	1	0	0
s. Another business education								
Required	2	0	0	3	3	4	1	1
Elective	98	100	100	97	97	95	99	99
Not offered	0	0	0	1	0	1	0	0
t. Agriculture								
Required	0	0	1	0	0	1	0	0
Elective	61	25	76	61	72	64	53	71
Not offered	39	75	23	39	28	36	47	29
u. Industrial arts								
Required	2	1	0	1	1	3	1	1
Elective	96	99	92	96	98	92	96	99
Not offered	3	0	8	3	1	4	3	0
v. Vocational education								
Required	2	0	1	3	2	3	2	2
Elective	93	91	93	93	98	87	95	98
Not offered	4	9	5	4	0	10	3	0
w. Work-study (e.g., DE)								
Required	1	0	0	2	0	1	1	2
Elective	92	97	92	88	94	79	96	98
Not offered	8	3	8	10	6	20	3	0
x. Driver education								
Required	19	12	16	20	21	18	18	21
Elective	77	77	80	78	76	79	77	73
Not offered	4	12	4	3	2	3	5	6

Instruction in Special Areas

- Of the specialized areas, only health education is more likely than not to be offered as a separate curriculum course in 63 percent of the school systems but integrated into other courses in 37 percent.

- Safety education is offered in every school system but is integrated into other courses in 83 percent of the school systems.

- Drug education, alcohol education, metric education, consumer education, and career education are offered by 99 percent of the school systems. More than nine-tenths, however, integrate drug education, alcohol education, and metric education into other courses. One-third (34 percent) offer consumer education and one-fifth (20 percent) offer career education as a separate course.

- More than 90 percent offer sex education, environmental studies, and human relations studies. The vast majority (i.e., over 80 percent) integrate these into other courses.

- Courses in international studies are not offered by 17 percent; the majority (71 percent) integrate such studies into other courses.

- A little more than half (57 percent) offer bilingual education--34 percent as a separate curriculum course and 23 percent as integrated into other courses. While only 33 percent of the small school systems offer bilingual education, 90 percent of those with 25,000 or more students offer this.

INSTRUCTION IN SPECIAL AREAS

		Region				Stratum		
Question	Total	Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
In general, does your school provide instruction in each of the following areas?								
a. Drug education								
Not offered	1	0	0	2	0	1	1	0
Integrated into other courses	92	80	95	93	94	92	91	92
Separate curriculum course	8	20	5	5	5	6	9	9
b. Sex education								
Not offered	8	9	9	7	6	7	8	8
Integrated into other courses	82	69	82	87	82	84	81	80
Separate curriculum course	11	23	9	6	12	9	11	13
c. Alcohol education								
Not offered	1	0	0	2	0	1	1	0
Integrated into other courses	93	83	96	93	97	93	92	93
Separate curriculum course	6	17	4	5	3	5	7	7
d. Bilingual education								
Not offered	43	58	42	53	33	67	46	10
Integrated into other courses	23	11	28	20	26	19	22	29
Separate curriculum course	34	31	31	27	41	14	33	61

INSTRUCTION IN SPECIAL AREAS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
e. Metric education								
Not offered	1	0	0	2	0	2	1	0
Integrated into other courses	96	98	96	95	96	95	97	95
Separate curriculum course	3	3	4	3	4	3	2	5
f. Health education								
Not offered	0	1	0	1	0	0	1	0
Integrated into other courses	37	37	41	32	31	37	35	40
Separate curriculum course	63	62	59	67	69	63	65	60
g. Safety education								
Not offered	0	0	0	1	0	0	1	0
Integrated into other courses	83	79	83	84	80	80	81	88
Separate curriculum course	17	21	17	15	20	20	18	12
h. Consumer education								
Not offered	1	0	4	1	0	1	3	0
Integrated into other courses	65	67	65	63	69	63	64	70
Separate curriculum course	34	33	31	36	31	37	33	30
i. Career education								
Not offered	1	0	1	1	1	0	2	0
Integrated into other courses	79	84	72	81	81	79	82	77
Separate curriculum course	20	16	27	18	18	21	16	23
j. International studies								
Not offered	17	9	19	18	16	24	14	11
Integrated into other courses	71	74	64	72	78	66	73	76
Separate curriculum course	12	17	16	10	6	10	13	14
k. Environmental studies								
Not offered	3	1	1	4	5	3	4	2
Integrated into other courses	87	82	86	87	90	86	87	90
Separate curriculum course	10	17	13	9	5	12	10	9
l. Human relations studies								
Not offered	5	4	5	7	4	7	4	3
Integrated into other courses	89	88	90	85	91	87	90	88
Separate curriculum course	7	9	5	8	5	5	6	9

Student Evaluation

- About half (51 percent) of the school systems require minimal competency testing of students, 35 percent for grade promotion and/or for graduation from high school and 17 percent for other reasons. School systems in the Southeastern region and those with 3,000 or more students are more likely to require such testing than are other school systems.

- Almost three-fifths (59 percent) provide the opportunity for early graduation from high school, with 2 percent actively promoting it. About three-tenths (29 percent) allow it only under unusual circumstances, whereas 13 percent do not allow it.

- When elementary students do not successfully complete the requirements for a grade, about two-fifths (38 percent) retain them in that grade until they do so. More than half (53 percent) promote them to the next grade, with most of those giving specially designed remedial help. Only 9 percent assign them to specially designed programs.

- The majority require the use of GROUP STANDARDIZED achievement tests (88 percent) and reading readiness tests (67 percent). About half require the use of intelligence tests (52 percent) and aptitude tests (46 percent). About one-fourth (28 percent) require the use of interest tests. Large school systems are slightly more likely to require achievement tests but less likely to require the other four than are small school systems.

- No published GROUP STANDARDIZED achievement test is used by a majority of the school systems, although the California Achievement Tests, the Comprehensive Test of Basic Skills, the Iowa Test of Basic Skills, and the Metropolitan Achievement Tests are required by a little more than one-third of the school systems. About one-fourth require the SRA Achievement Series and the Stanford Achievement Tests. Some school systems require the use of two or more. The PSAT, SAT, and ACT are generaly made available but not required.

- More than three-fourths (78 percent) of the school systems report using criterion-referenced achievement tests. These are used less often in the Northeast but more often in large school systems.

- About two-fifths (43 percent) place some reliance on GROUP STANDARDIZED achievement test scores as reliable measures of the effectiveness of individual schools.

- In almost two-fifths (37 percent) of the school systems, the mean of GROUP STANDARDIZED achievement test scores for each regular school is reported to the media. This is practiced more often in school systems in the Southeastern region and in large school systems than in others.

- The majority use scores from GROUP STANDARDIZED achievement tests at the building level for diagnosing individual student learning needs (87 percent) and evaluating the curriculum (75 percent). A significant number use the scores for tracking or grouping students (43 percent) and evaluating effectiveness as compared with that of other schools (32 percent). Few use them for determining promotion/retention of students (9 percent) or evaluating individual teacher effectiveness (5 percent).

STUDENT EVALUATION

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Does your school system require MINIMAL COMPETENCY TESTING of students? (Percent responding.)								
Yes, for grade promotion	4	3	4	5	3	6	1	4
Yes, for graduation from high school	25	24	41	13	23	9	31	34
Yes, for both grade promotion and graduation from high school	6	5	14	3	8	4	6	10
Neither of those, but for other reasons	17	19	14	19	16	14	17	21
No	49	51	28	61	50	67	46	30
Which of the following BEST describes your school system's policy on EARLY GRADUATION (e.g., when a student meets graduation requirements at the end of the eleventh grade)? (Percent responding.)								
Early graduation is actively promoted	2	1	4	1	3	1	1	4
All students are given the opportunity to graduate early but it is NOT actively promoted	57	47	55	58	55	46	58	68
Early graduation is allowed only under unusual circumstances	29	34	34	29	26	30	33	22
Early graduation is NOT allowed	13	17	8	13	16	23	8	7
When elementary students DO NOT successfully complete the requirements for a grade, which of the following BEST describes the policy of your school system? (Percent responding.)								
Retain them in grade until they meet the requirements	38	42	41	40	36	43	33	37
Take them out of regular program and assign them to specially designed programs	9	11	13	8	8	9	8	12
Promote them to next grade but give them specially designed remedial help	49	47	43	47	52	45	56	45
Promote them to next grade	4	0	3	5	5	3	3	6

STUDENT EVALUATION—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Listed below are some areas for which GROUP STANDARDIZED TESTS are available. For which of these areas does your school system *require* (either by state mandate or local option) the use of the tests, for which are tests made available but their use is *optional*, and for which are *no* tests made available? (Percent responding.)								
a. Intelligence								
Required	52	57	62	63	35	53	56	43
Optional	29	29	25	25	34	37	27	22
Not used	19	14	13	12	31	11	17	35
b. Achievement								
Required	88	84	96	88	85	80	89	97
Optional	11	15	4	12	11	18	10	2
Not used	1	1	0	1	1	2	1	1
c. Reading readiness								
Required	67	68	72	71	61	67	74	59
Optional	28	27	25	28	26	28	22	34
Not used	5	5	3	1	13	4	5	7
d. Aptitude								
Required	46	49	56	44	42	48	49	38
Optional	45	39	37	51	43	42	44	50
Not used	10	12	7	5	15	10	7	12
e. Interest								
Required	28	27	35	33	21	36	28	18
Optional	61	57	51	64	61	51	63	72
Not used	11	16	14	3	19	13	9	11
Specifically, which of the following GROUP STANDARDIZED ACHIEVEMENT TESTS does your school system *require*, which are *optional*, and which are *not used*? (Percent responding.)								
a. California Achievement Tests								
Required	40	27	39	42	47	37	38	43
Optional	19	15	18	28	16	25	14	18
Not used	42	58	44	30	38	38	47	39
b. Comprehensive Tests of Basic Skills								
Required	39	34	50	27	38	29	43	44
Optional	17	15	16	23	19	18	19	14
Not used	44	52	34	50	44	53	38	42

STUDENT EVALUATION—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
c. Iowa Test of Basic Skills								
Required	35	34	18	44	30	42	33	30
Optional	19	16	26	24	16	25	16	17
Not used	46	50	56	32	54	34	52	53
d. Iowa Test of Educational Development								
Required	17	4	18	23	20	24	15	11
Optional	15	11	16	17	20	23	11	12
Not used	68	85	67	61	61	53	74	77
e. SRA Achievement Series								
Required	26	19	38	26	21	33	25	17
Optional	27	22	25	31	27	30	30	21
Not used	47	59	38	44	52	37	45	62
f. Stanford Achievement Tests								
Required	29	31	26	28	38	38	28	22
Optional	28	31	17	36	28	34	24	27
Not used	43	39	57	36	35	28	49	52
g. Metropolitan Achievement Tests								
Required	34	47	39	33	29	31	40	27
Optional	26	24	24	33	17	31	23	24
Not used	41	29	37	35	54	38	36	49
h. PSAT								
Required	11	11	10	13	7	16	13	2
Optional	78	78	75	81	75	70	79	86
Not used	11	11	15	6	19	14	9	12
i. SAT								
Required	13	12	11	13	9	16	17	2
Optional	81	80	82	85	77	75	79	90
Not used	7	8	7	2	14	10	4	8
j. ACT								
Required	9	5	3	11	10	12	12	2
Optional	76	62	80	79	77	75	72	82
Not used	16	33	17	10	13	13	16	17
Does your school system use any criterion-referenced achievement tests (these identify specific learning tasks students can perform) in all or most schools? (Percent responding.)								
Yes, required by the state	25	10	40	27	25	11	25	40
Yes, but only recommended by the state	3	3	3	3	2	5	2	2
Yes, but it was a local option	50	53	42	47	50	49	51	49
No	22	35	16	23	23	34	22	9

STUDENT EVALUATION—(CONTINUED)

		Region				Stratum		
Question	Total	Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Does your school system rely on GROUP STANDARDIZED ACHIEVEMENT TEST SCORES as reliable measures of the effectiveness of individual schools? (Percent responding.)								
Yes, definitely	4	0	3	3	11	2	4	8
Yes, somewhat	39	37	47	41	35	41	38	38
No	57	63	51	56	54	57	58	55
Does your school system report the *mean* of GROUP STANDARDIZED ACHIEVEMENT TEST SCORES for each *regular* school to the media? (Percent responding.)								
Yes	37	31	55	26	42	26	34	55
No	63	69	45	74	58	74	66	46
For the most part, how are *scores* from GROUP STANDARDIZED ACHIEVEMENT TESTS used at the school building level? (Percent responding. Percents may add to more than 100.)								
The tests are not used in this school system	2	5	0	2	2	2	2	1
Scores are not used at the building level	3	3	1	1	6	2	3	2
Diagnosing individual student learning needs	87	86	92	90	84	86	88	87
Tracking or grouping students	43	55	37	48	37	38	51	39
Determining promotion/retention of students	9	9	9	8	13	7	10	10
Evaluating individual teacher effectiveness	5	3	10	5	4	8	4	3
Evaluating the curriculum	75	78	77	69	79	63	81	80
Evaluating effectiveness as compared with that in other schools	32	29	32	30	40	32	32	30

Student Discipline

- Only 7 percent report that student discipline is NO problem in their school system. Although few (3 percent) say it is a major problem, about two-fifths (39 percent) say it is a moderate problem. Almost two-thirds (66 percent) of the larger school systems, however, report discipline as at least a moderate problem, with 8 percent calling it a major problem.

- On the average, 7 percent of the individual schools in the school systems have SERIOUS student discipline problems.

- Although a wide variety of practices to help schools and teachers with their discipline problems are mentioned, about half of the school systems surveyed say they have "provided more extensive vocational programs" (51 percent) and "provided special staff training on how to handle students misbehavior" (50 percent). About 40 percent report "providing more work-study opportunities," "writing strict discipline codes," and "allowing limited corporal punishment." In addition, about two-thirds of the large school systems provide "special schools or classes for students who are chronic problems."

STUDENT DISCIPLINE

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
In general, to what extent is student discipline a problem in your school system? (Percent responding.)								
Major problem	3	3	5	1	5	0	1	8
Moderate problem	39	26	59	36	36	25	39	58
Small problem	51	59	31	56	52	60	56	33
No problem	7	12	5	7	7	15	5	2
What percent of your schools would you estimate as having SERIOUS student discipline problems?								
5 percent or less	63	59	46	66	72	81	54	57
6 to 14 percent	24	26	26	31	16	13	29	28
15 to 29 percent	9	15	15	3	6	3	12	12
30 to 49 percent	3	0	9	0	6	0	5	4
50 percent or more	1	0	4	0	0	3	0	0
Mean percent	7	6	12	5	7	6	8	8
Median percent	5	5	10	5	5	3	5	5

STUDENT DISCIPLINE—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
What has your school system done to help schools and teachers with their student discipline problems? (Percent responding. Percents may add to more than 100.)								
Nothing specific	7	8	1	8	8	15	3	2
Provided more work-study opportunities	44	55	47	46	42	29	50	55
Provided more extensive vocational programs	51	44	63	55	49	35	59	61
Required at least some schools to provide more traditional curricula and classroom organization	15	13	18	13	14	8	15	23
Provided special staff training on how to handle student misbehavior	50	38	54	49	52	29	50	77
Hired more paraprofessionals to work with teachers	31	26	30	31	34	24	33	38
Made parents more responsible for the discipline of their children	35	38	38	32	34	37	31	37
Hired more counselors and other specialists to work with problem students	27	29	22	28	24	15	27	41
Reduced class sizes in problem schools	14	13	11	13	15	12	11	23
Written strict discipline codes	42	46	51	38	46	35	43	49
Hired security personnel in problem schools	18	19	20	12	26	2	14	45
Given teachers more authority to handle discipline problems	25	23	18	19	29	29	21	25
Provided special school(s) for students who are chronic problems	31	26	37	26	34	8	28	66
Provided for special classes in some schools for students who are chronic problems	38	31	43	38	39	17	38	65
Allowed limited corporal punishment	39	15	49	43	48	36	39	45
Established broader provisions for suspension	32	26	34	36	32	34	30	33
Established broader provisions for explusion	21	14	28	22	24	16	21	27

Financing

- The median budget is $10 million with a low of $127 thousand and a high of almost $3 billion.
- The majority (86 percent) report an increase in their 1979-80 budget. The median increase was 8 percent, with only 6 percent reporting a 3 percent decrease on the average.
- For 1979-80, 15 percent report cutbacks in some programs/services and 29 percent report expanding some programs/services. A slight majority (56 percent) have maintained the status quo.
- When cutbacks are made, they are more likely to be made in numbers of teachers (59 percent) or numbers of administrative/supervisory personnel (52 percent). Forty-two percent have cut back on maintenance/custodial services. Few have cut back on kindergarten and/or other preschool programs (5 percent) or food services (7 percent).
- One-third (33 percent) consider the funding of their current program adequate, another one-third (33 percent) barely adequate, more than one-fourth (28 percent) inadequate, and (6 percent) very inadequate.
- To enable the school systems to do a job in 1979-80 that they felt was average would generally have required a median increase of 20 percent over the 1978-79 budget.
- Three-fifths (61 percent) report the funding of their 1979-80 budget as a problem which they expect to continue for the next three years. Another 27 percent say that although it is not now a problem, they expect it to be one in the next three years. Only 12 percent foresee no immediate problem.
- Four-fifths (81 percent) feel funding of education is a problem for at least three-fourths of the school systems in their state.
- Only 3 percent blame inadequate amounts of federal aid as the major difficulty in maintaining adequate school funding. More than half (53 percent) say the major problem is movements to "hold the line on property taxes," and 44 percent say it is "inadequate amounts of state aid."
- More than one-fourth (27 percent) report successful reductions or limitations on property taxes in the community where their school system is located. Another 13 percent report unsuccessful attempts. About one-third (34 percent) say that although it hasn't happened yet, it's likely to happen. Only 27 percent foresee no such attempts.
- The mean percentage of current expenditure revenues provided by the three levels of government are local: 44 percent; state: 49 percent; and federal: 7 percent. About one-tenth (9 percent) say more than 75 percent comes from local sources, and 7 percent say more than 75 percent comes from state sources.
- The majority (57 percent) oppose the contention that the federal government's share of the cost of public education must be at least one-third.
- Nine-tenths (91 percent) report receiving DIRECT federal aid through ESEA's Education for the Economically/Educationally/Disadvantaged Children, 73 percent through Education for All Handicapped Children, 49 percent through impact aid, and 12 percent through Teacher Centers and/or Teacher Corps programs.

- Half (52 percent) prefer that federal aid be distributed generally among all school systems, while one-sixth (15 percent) prefer it be distributed categorically. Almost one-third (31 percent) prefer a combination of general and categorical aid.
- Only about one-third (36 percent) of the school systems are fiscally independent, with more than three-fifths (62 percent) being fiscally dependent. School systems in the Southeastern region are more likely to be fiscally dependent than others.
- To meet the increased energy costs for busing of students, about one-third (34 percent) cut back on the use of energy. Another two-fifths met the energy costs by finding new sources of revenue (16 percent) or reducing services in other areas (23 percent).
- Two-thirds (65 percent) cut back on the use of energy for heating/air conditioning, whereas most of the remaining school systems (30 percent) met the increased costs.

FINANCING

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
What is your 1979-80 school year budget (in thousands of dollars) for current expenditures? (DO NOT include capital outlay or interest on building bonds.) (Percent responding.)								
Less than $1,000,000	6	4	3	9	4	18	0	0
$1,000,000 to $9,999,999	44	59	37	48	34	77	46	0
$10,000,000 to $49,999,99	24	22	24	23	25	1	49	17
$50,000,000 or more	26	15	36	21	36	4	5	83
Median budget (Millions of dollars)	$10M	$7.5M	$16M	$7M	$17.6M	$2M	$11M	$75M
What percentage change was this over your 1978-79 budget? (Percent responding.)								
Increased	86	90	84	87	87	84	88	86
Median increase (percent)	8	6	8	8	9	9	7	7
Relatively no change	8	9	10	9	7	10	6	10
Decreased	6	1	6	5	6	6	6	4
Median decrease (percent)	3	3	4	5	0	7	2	2
In general, to what extent were you able to maintain instructional services for 1979-80? (Percent responding.)								
Cut back some services/programs	15	16	11	10	19	9	17	20
Cut back some services and expanded some services, but generally maintained the status quo	17	17	23	15	15	8	16	27

FINANCING—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Little or no change in any services/programs	39	37	31	49	36	53	34	31
Expanded some services/programs	29	30	35	26	30	29	34	22
If CUTBACKS were made (1 or 2 checked above), in what areas did they occur? (Percent responding who had "cutbacks." Percents may add to more than 100.)								
Class sizes	18	19	12	17	13	14	21	15
Numbers of teachers	59	73	50	74	53	57	63	54
Numbers of professional instructional support personnel (e.g., counselors)	30	12	23	31	41	14	27	41
Numbers of supervisory/administrative personnel	52	27	58	40	75	14	47	75
Numbers of nonprofessional instructional support personnel (e.g., teacher aides)	33	23	35	29	47	21	26	46
Maintenance/custodial services	42	19	50	34	63	18	44	53
Busing/transportation services	19	15	15	23	16	32	21	12
Food services	7	8	4	9	9	7	10	5
Kindergarten or other preschool programs	5	0	15	0	3	4	2	9
Some curricular programs (1-12) (e.g., music, art)	34	39	35	29	34	36	39	29
Extracurricular activities	19	12	23	14	13	18	23	17
Sports programs	23	19	27	14	22	18	24	25
School construction/remodeling	27	12	31	26	38	21	19	37
Regardless of whether or not you have made cutbacks in services, do you consider the funding of your current program adequate or inadequate for you to provide a good education for your students? (Percent responding.)								
Funding is adequate	33	42	25	26	39	39	34	23
Funding is barely adequate	33	32	35	39	28	36	32	32
Funding is inadequate	28	22	33	29	26	21	29	37
Funding is VERY inadequate	6	4	6	6	7	5	6	8
What percentage increase in your 1978-79 budget would have enabled you to do a job you think would be average?								
Median percent increase	20	23	20	15	20	20	20	15

FINANCING—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
To what extent is the funding of your budget going to be a problem in the next three years? (Percent responding.)								
It is NOT now a problem and shouldn't be one in the next three years	9	6	12	7	12	14	6	6
It is NOT now a problem but will be one within three years	27	20	14	35	31	37	28	15
It is now a problem but should get better within three years	3	3	4	5	2	4	3	2
It is now a problem and probably will continue to be one in the next three years	61	72	71	53	55	46	63	76
To what extent does the funding of education seem to be a problem in *your state*? (Percent responding.)								
Not a problem for any school system in this state	1	3	1	0	1	1	1	2
Only a problem for a *few* school systems in this state	7	5	8	9	6	8	6	10
A problem for about *half* of the school systems in this state	10	8	7	12	16	14	7	12
A problem for about *three-fourths* of the school systems	12	13	8	13	15	17	11	5
A problem for *most* school systems .	69	73	76	66	62	61	76	71
As far as your school system is concerned, what is the major reason that keeping the schools adequately funded is difficult? (Percent responding.)								
Movements to "hold the line" on local property tax increases	53	48	69	56	51	47	56	56
The inadequate amount of state aid to local school systems	44	51	28	42	45	48	44	40
The inadequate amount of federal aid	3	1	3	3	4	5	1	4
Has there been any organized attempt in the community where your school system is located to reduce or limit property taxes?								
Yes, it has happened successfully . . .	27	30	24	14	35	14	29	39
Yes, it has happened but NOT successfully	13	15	15	14	9	8	13	18
No, but it is likely to happen	34	40	36	39	26	43	35	21
No, and it is NOT likely to happen .	27	15	24	33	29	35	23	22

FINANCING—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
What percentage of your current expenditure revenues comes from local, state, and federal sources? (Round off to the nearest whole percent.) (Percent responding.)								
Local								
25 percent or less	23	10	27	15	30	18	26	24
26 to 50 percent	39	24	44	41	43	40	37	43
51 to 75 percent	28	43	25	32	21	28	28	28
76 percent or more	9	24	4	12	6	14	9	4
Mean percent	44	59	39	48	40	48	43	42
Median percent	45	60	35	49	41	48	43	41
State								
25 percent or less	13	34	4	15	9	20	12	6
26 to 50 percent	39	44	35	43	38	40	37	41
51 to 75 percent	41	20	60	39	41	36	44	44
76 percent or more	7	3	1	3	12	5	7	9
Mean percent	49	37	52	47	53	46	50	52
Median percent	50	38	56	48	53	45	51	52
Federal								
Mean percent	7	5	10	6	7	6	7	7
Median percent	5	4	9	5	5	5	5	6
Some people contend that the federal government's share of the cost of public education must be at least one-third. Would you support or oppose that contention? (Percent responding.)								
Strongly support	17	28	23	14	12	18	15	21
Tend to support	26	29	37	25	19	23	28	25
Tend to oppose	29	25	20	35	26	27	32	28
Strongly oppose	28	19	20	27	43	32	25	26
Does your school receive any DIRECT aid through any of the following federal aid programs? (Percent responding "yes.")								
Impact Aid (for children of parents who live or work on federal property)	49	43	52	39	61	29	40	88
ESEA: Education for the Economically/Educationally Disadvantaged Children	91	91	94	89	92	91	87	96

FINANCING—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Education for All Handicapped Children	73	57	80	69	81	53	79	88
Teacher Centers and/or Teacher Corps	12	7	19	7	16	1	6	34
Does your school system prefer that federal aid be distributed CATEGORICALLY as it is now, or that it be prorated GENERALLY among all school systems?								
Categorical aid	15	19	12	14	14	16	16	12
General aid	52	44	58	53	49	56	50	48
A combination of categorical and general aid	31	32	28	32	36	23	33	39
Neither	2	5	3	1	1	5	1	1
Is your school system FISCALLY INDEPENDENT or is it FISCALLY DEPENDENT upon another level of government for funding? (Percent responding.)								
Fiscally independent	36	48	18	42	31	39	38	31
Fiscally dependent	62	51	78	57	67	60	61	65
Other	2	1	4	1	2	1	1	4
Busing of students and heating/air conditioning the schools are two big uses of energy by the public schools. What has your school system done in each case to meet the increased costs? (Percent responding.)								
Busing								
a. Cut back the use of energy	34	37	25	36	33	42	33	27
b. Found *new* sources of revenue to meet increased costs	16	7	24	16	25	11	19	20
c. Reduced services in other areas to meet the increased costs	23	19	19	28	22	23	24	21
d. None of these	27	37	33	21	20	25	25	32
Heating/air conditioning								
a. Cut back the use of energy	65	73	61	59	68	64	68	61
b. Found *new* sources of revenue to meet increased costs	9	4	13	10	14	7	10	11
c. Reduced services in other areas to meet the increased costs	21	16	21	27	15	20	20	24
d. None of these	5	7	5	4	3	9	3	3

Censorship

- About one-sixth (16 percent) of the school systems have faced attempts at censorship over the last two years; however, only one-seventh of those attempts (2 percent of the total) were successful. Attempts at censorship occurred less often in school systems in the Northeastern region and in small school systems.
- When attempts were made, in a majority of the cases textbooks and library books were the main targets, sex and pornography were the area of attack, and individuals in the community acting on their own were responsible.

CENSORSHIP

		Region				Stratum		
Question	Total	Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Has your school system faced any attempts at censorship in the last two years (1978-79 and 1979-80)?								
Yes, some successful attempts	2	0	0	2	1	2	1	2
Yes, some serious threats but with little success	14	10	18	11	23	5	13	28
No	84	90	82	83	76	93	86	70
If YES, answer each of the following:								
• Who was primarily responsible for the censorship attempts? (Percent responding who said "yes" above.)								
Individuals in the community acting on their own	71	100	57	68	65	92	79	59
Organized group(s) from the community	22	0	36	23	22	8	18	28
Organized group(s) in the state targeting this community	4	0	7	0	9	0	4	5
Organized national group(s) targeting this community	4	0	0	9	4	0	0	8
• What were the targets? (Percent responding who said "yes" above. Percents may add to more than 100.)								
Textbooks	56	57	57	52	63	67	48	59
Library books	59	57	64	81	33	83	62	49
Magazines, etc.	11	11	14	19	0	33	10	5
Specific curriculum courses or areas	40	22	50	24	50	33	38	44
Specific teachers or administrators	9	0	14	10	4	25	7	5
Other	9	11	0	0	17	0	10	10

CENSORSHIP—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
• What were the general area(s) of attack? (Percent responding who said "yes" above. Percents may add to more than 100.)								
Sex and/or pornography	66	78	78	73	46	83	75	55
Drugs/tobacco/alcohol	11	11	7	9	17	33	14	3
Religion versus science	34	33	71	18	33	33	39	30
Political and/or economic systems	15	11	29	9	21	25	14	13
Race and/or ethnicity	15	0	14	23	13	0	7	25
Sex role stereotyping	10	0	0	9	21	8	4	15
Other	25	22	7	14	38	17	25	28

Handicapped Students

- In over half (55 percent) of the school systems, handicapped students attend the school in their attendance area. One-fifth (20 percent) bus them to selected schools and one-tenth (11 percent) have specifically designed centers for handicapped students. About one-seventh (14 percent) have no dominant practice. School systems in the Southeast are slightly more likely to send them to schools in their attendance areas.

- Aside from a few exceptions (5 percent or less), most school systems provide for the instruction of students who are speech impaired, deaf or hard of hearing, visually handicapped, mentally retarded, seriously emotionally disturbed, orthopedically handicapped or other health impaired.

- Students who are speech impaired are more commonly mainstreamed for ALL instruction while students with the other handicaps are more likely to be mainstreamed only for some instruction. Students who are seriously emotionally disturbed are more likely than those with other handicaps to have ALL instruction in special classes.

- The estimates of percentages of students in the school systems who have the kinds of learning disabilities defined in P.L. 94-142 averaged a mean of 8 percent.

- More than nine-tenths of the school systems provide the following for all or most handicapped students: established procedures to assure early identification (98 percent); individual education programs (95 percent); all education in a "least restrictive environment" (94 percent); and scheduled time with specialists in their handicap (94 percent). About four-fifths (81 percent) provide transportation to and from school for all or most handicapped students, although only 69 percent of the large school systems provide such transportation.

- In general, special provisions to help teachers who have handicapped students are available more often for the special education teacher than for the regular classroom teacher. A large majority of school systems provide information on P.L. 94-142, in-service education, procedures to request instructional materials, help from specialists, procedures for review of placement, involvement in IEP's, information on due process rights of parents, and help from special teachers for both special education and regular classroom teachers. The three exceptions are as follows:

  a. Two-thirds (66 percent) decrease the average student load for special education teachers, but only 12 percent for regular classroom teachers.

  b. Only two-fifths (39 percent) provide a teacher aide or parent volunteer for a special education teacher when a student is mainstreamed, and only 28 percent for regular classroom teachers.

  c. Only one-third (34 percent) provide released or compensatory time for special education teachers, but only 13 percent for regular classroom teachers.

HANDICAPPED STUDENTS (P.L. 94-142)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
In general, how does your school system handle the education of handicapped students? (Percent responding.)								
They attend the school in their attendance area	55	54	65	47	55	60	52	53
Most are bused to selected schools	20	21	15	25	18	15	26	19
They attend specifically designed centers for the handicapped	11	10	9	14	10	12	11	9
No dominant practice	14	15	12	15	17	14	11	19
P.L. 94-142 tries to assure that each handicapped child is placed in REGULAR classes (mainstreamed) when it's appropriate. In general, for how much of their instruction are children who have the following kinds of handicaps mainstreamed in your school system? (Percent responding.)								
a. Speech impaired								
Elementary								
No provision for instruction	0	0	0	0	0	0	0	0
All instruction in special classes	4	3	3	3	3	4	4	2
Mainstreamed for some instruction	34	32	41	32	36	42	26	36
Mainstreamed for all instruction	62	66	56	65	61	54	70	63
Secondary								
No provision for instruction	1	0	0	1	2	2	1	0
All instruction in special classes	3	3	1	3	4	5	4	1
Mainstreamed for some instruction	31	28	39	30	30	37	26	32
Mainstreamed for all instruction	65	69	60	67	64	56	70	67
b. Deaf or hard of hearing								
Elementary								
No provision for instruction	5	3	1	7	7	10	5	0
All instruction in special classes	15	18	18	16	14	18	16	10
Mainstreamed for some instruction	62	49	65	62	71	49	57	85
Mainstreamed for all instruction	18	31	15	15	8	23	22	5

HANDICAPPED STUDENTS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Secondary								
No provision for instruction ...	5	1	1	7	8	11	5	0
All instruction in special classes ...	12	14	17	13	9	13	13	9
Mainstreamed for some instruction ...	61	44	65	60	73	47	57	81
Mainstreamed for all instruction ...	22	40	17	21	10	30	24	11
c. Visually handicapped								
Elementary								
No provision for instruction ...	5	1	1	7	7	8	6	0
All instruction in special classes ...	12	14	14	13	12	14	13	9
Mainstreamed for some instruction ...	57	51	58	56	61	46	54	73
Mainstreamed for all instruction ...	26	34	27	24	21	32	26	18
Secondary								
No provision for instruction ...	5	1	1	7	7	8	7	0
All instruction in special classes ...	9	10	13	10	10	11	8	8
Mainstreamed for some instruction ...	54	48	54	56	57	43	65	65
Mainstreamed for all instruction ...	31	41	31	27	26	39	28	28
d. Mentally retarded								
Elementary								
No provision for instruction ...	0	0	0	1	1	1	0	0
All instruction in special classes ...	17	14	16	15	19	20	17	16
Mainstreamed for some instruction ...	79	84	84	81	76	74	82	81
Mainstreamed for all instruction ...	3	1	0	4	4	5	1	3
Secondary								
No provision for instruction ...	0	0	0	1	1	1	0	0
All instruction in special classes ...	15	14	12	13	17	16	14	15
Mainstreamed for some instruction ...	81	85	88	81	77	76	85	82
Mainstreamed for all instruction ...	4	1	0	5	4	7	1	4

HANDICAPPED STUDENTS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
e. Seriously emotionally disturbed								
Elementary								
No provision for instruction ...	3	1	1	5	5	7	3	0
All instruction in special classes	36	43	48	38	28	31	38	39
Mainstreamed for some instruction	57	54	44	54	65	54	59	57
Mainstreamed for all instruction	4	1	7	3	2	8	1	2
Secondary								
No provision for instruction ...	4	3	1	5	7	7	4	1
All instruction in special classes	34	38	44	36	28	28	37	35
Mainstreamed for some instruction	59	58	49	55	63	56	58	62
Mainstreamed for all instruction	4	1	6	4	3	8	1	2
f. Orthopedically impaired or other health impaired								
Elementary								
No provision for instruction ...	2	0	0	4	2	6	1	0
All instruction in special classes	16	16	19	17	17	10	21	17
Mainstreamed for some instruction	59	49	51	63	64	55	55	71
Mainstreamed for all instruction	22	35	30	17	17	30	23	12
Secondary								
No provision for instruction ...	3	0	0	4	3	5	3	0
All instruction in special classes	13	13	16	16	14	9	17	13
Mainstreamed for some instruction	60	53	50	62	65	55	55	73
Mainstreamed for all instruction	24	34	34	18	18	31	26	14
What do you estimate as the percent of students in your school system who have the kinds of learning disabilities defined in P.L. 94-142?								
Less than 5 percent	25	19	26	23	30	26	25	24
5 to 9 percent	35	33	36	37	38	35	36	32
10 to 14 percent	28	30	25	28	26	23	28	35
15 to 19 percent	8	12	7	8	3	10	7	6
20 percent or more	4	6	6	4	3	5	4	3
Mean percent	8	9	8	8	8	8	9	9
Median percent	8	9	7	8	7	8	8	8

HANDICAPPED STUDENTS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
In general, how many handicapped students are provided each of the following? (Percent responding.)								
a. Transportation to and from school								
For all or most	81	88	91	88	69	86	85	69
For about half	8	1	5	4	16	7	5	14
For more or only a few	11	11	4	9	15	8	10	17
b. All education in a "least restrictive environment"								
For all or most	94	95	87	97	93	94	93	94
For about half	4	4	8	2	4	2	5	4
For more or only a few	3	1	5	1	3	4	2	2
c. An Individualized Education Program								
For all or most	95	98	97	94	97	89	99	97
For about half	2	1	1	5	0	6	2	0
For more or only a few	3	1	1	2	3	6	0	3
d. Scheduled time with specialists in their handicap								
For all or most	94	99	94	94	92	88	98	95
For about half	3	0	4	2	2	5	1	2
For more or only a few	4	1	3	4	6	7	1	3
e. Procedures established to assure early identification								
For all or most	98	100	99	98	97	96	100	97
For about half	2	0	1	1	2	2	1	2
For more or only a few	1	0	0	1	1	1	0	1
What special provisions have been made to provide help for both *special education teachers* and *regular classroom teachers* who have handicapped students mainstreamed in their class(es)? (Percent responding "yes.")								
a. Information to ALL teachers on the implication of P.L. 94-142								
Special education teachers	97	97	100	98	97	97	97	98
Regular classroom teachers	86	80	88	85	92	89	85	84
b. In-service education on how to work with handicapped children								
Special education teachers	95	92	97	94	97	92	96	98
Regular classroom teachers	81	89	78	81	77	80	81	80

HANDICAPPED STUDENTS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
c. Procedures for *requesting* and *receiving* appropriate instructional materials and equipment								
Special education teachers	95	90	99	94	96	92	94	99
Regular classroom teachers	71	80	78	70	68	75	69	79
d. Procedures for *requesting* and *receiving* help from a school psychologist, counselor, or other specialist								
Special education teachers	97	93	99	96	99	96	97	99
Regular classroom teachers	94	91	93	93	98	90	94	96
e. Released or compensatory time for the extra work involved								
Special education teachers	34	34	26	36	33	38	34	30
Regular classroom teachers	13	18	5	14	7	14	15	8
f. Decrease from the average student load								
Special education teachers	66	58	69	69	70	62	74	59
Regular classroom teachers	12	11	9	15	12	11	12	12
g. Procedures for requesting review of placement of a handicapped student								
Special education teachers	94	90	96	92	97	90	95	96
Regular classroom teachers	87	86	93	86	88	82	89	89
h. Involvement in the preparation of Individualized Education Programs								
Special education teachers	96	97	96	97	97	92	97	99
Regular classroom teachers	75	70	75	77	78	74	77	74
i. Specific information provided to all teachers on the due process rights of parents								
Special education teachers	94	90	96	95	96	92	95	96
Regular classroom teachers	77	75	82	76	81	75	77	81
j. Help made available from special teachers whenever it's needed								
Special education teachers	83	86	85	81	80	81	83	84
Regular classroom teachers	78	81	76	77	77	74	79	81
k. A teacher aide or parent volunteer assigned when a handicapped student is mainstreamed								
Special education teachers	39	42	30	38	41	39	44	30
Regular classroom teachers	28	29	17	26	29	31	29	22

Professional Development

- For one-third (33 percent) of the school systems, the primary intent for the professional development of teachers is to provide group training that meets school system needs. Another 13 percent have the primary intent of providing opportunities that meet individual teacher needs. Half (50 percent) give both of these special attention.

- The most commonly used provision for the professional development of teachers in 1979-80 were workshops or special courses offered during the regular school year by 86 percent of the school systems. Provisions for university/college extension courses offered in the community were made by 67 percent of the school systems. About half provided for workshops or special courses offered during the summer (54 percent) and a specified number of days of released time for conferences, etc. (51 percent). Only about one-third provide tuition reimbursement (32 percent) and sabbatical leaves (36 percent).

- A majority of school systems (63 percent) REQUIRE, when appropriate, in-service education to prepare teachers for new curricula. About one-third require it to correct identified weaknesses in performance (39 percent), for state certification (35 percent), for advancement on the salary schedule (35 percent), and for state recertification (32 percent). Only about one-sixth require it to comply with negotiated agreements (18 percent) and to help fulfill degree requirements (17 percent). Eight percent require it to fulfill requirements for tenure.

PROFESSIONAL DEVELOPMENT

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Which of the following best describes your school system's *primary* intent for the professional development of teachers? (Percent responding.)								
To provide group training that meets school system needs (e.g., preparation for new curricula, training in classroom management techniques)	33	21	30	33	39	37	31	32
To provide opportunities for professional growth that meet individual teacher needs (e.g., supporting advanced degree programs, released time for individual activities)	13	11	14	11	14	17	15	5
Both of these get equal attention	50	64	51	51	46	41	51	61
Neither of these	0	0	1	0	0	0	1	0
Don't have any programs	3	4	4	4	1	5	2	2

PROFESSIONAL DEVELOPMENT—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
What specific provisions are being made by your school system in 1979-80 for the professional development of teachers? (Percent responding. Percents may add to more than 100.)								
None	2	1	1	3	2	5	2	0
University/college extension courses offered in the community	67	53	83	67	68	54	73	73
Some or all tuition reimbursed for study at colleges or universities	32	57	40	30	18	27	37	29
Workshops or special courses offered during the regular school year	86	89	94	80	86	74	89	95
Workshops or special courses offered during the summer	54	40	72	49	61	28	60	80
Staffed Teacher Center(s)	17	9	22	17	23	7	10	39
Unstaffed Teacher Center(s)	5	3	8	4	4	5	4	6
A specified number of sabbatical leaves, either for a year or a semester	36	58	28	29	35	18	41	51
A specified number of days of released time for each teacher to attend conferences, visit other schools/programs, etc.	51	63	44	48	51	53	54	47
Short periods of released time available for individual study	22	17	32	20	21	18	24	24
A dollar allocation for each teacher to use for his/her own professional development	4	4	1	6	3	5	4	2
Other	9	12	12	7	7	6	9	12
Is in-service education REQUIRED of teachers, whenever appropriate, for any of the following reasons? (Percent responding. Percents may add to more than 100.)								
To help fulfill degree requirements	17	19	28	14	11	18	15	17
For state certification	35	39	49	31	35	37	31	37
For state recertification	32	19	53	38	22	30	26	41
For advancement on the salary schedule	35	44	28	31	34	35	35	34
To fulfill requirements for tenure	8	15	6	7	8	6	5	16

PROFESSIONAL DEVELOPMENT—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
To correct identified weaknesses in performance	39	26	54	39	34	35	41	40
To prepare teachers for new curricula	63	62	64	59	66	56	67	67
To comply with negotiated agreements	18	25	8	20	15	17	17	19
Other	9	3	7	6	20	6	8	17

Teacher Evaluation

- Almost nine-tenths of the school systems (88 percent) DO NOT require teachers to pass a competency test before they are certified to teach. However, 12 percent have such a requirement for first certification only, with almost one-third (33 percent) of the school systems in the Southeast maintaining that requirement.
- Of those school systems requiring teachers to pass a competency test, most (94 percent) do so because of a state requirement.
- Where the tests are used, about half of the school systems (48 percent) said the competency test was developed by the state, and a little less than half (44 percent) said it was commercially developed.
- About half (53 percent) evaluate ALL professional staff at least once a year by means other than a competency test. One-sixth (17 percent) evaluate only classroom teachers, with nearly another one-fifth (19 percent) evaluating only some classroom teachers (e.g., nontenured). Eleven percent evaluate all building-level professional staff, including principals.
- Evaluations other than competency tests are used by school systems mainly for reappointment/dismissal of teachers (72 percent), selecting teachers for tenure (51 percent), and selecting teachers for administrative/supervisory positions (39 percent). Only 14 percent use other means of evaluation for advancing teachers on the salary schedule and 4 percent for awarding merit pay.

TEACHER EVALUATION

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Are classroom teachers in your school system *required* to pass a COMPETENCY TEST before they are certified (licensed) to teach? (Percent responding.)								
Yes, for first certification only	12	3	33	16	2	8	11	20
Yes, for recertification only	0	0	0	0	0	0	0	0
Yes, for both first certification and recertification	0	0	1	0	0	0	1	0
No	88	98	65	84	98	92	89	80
If YES, answer each of the following:								
• Is it a state requirement or your school system's requirement? (Percent responding who said "yes" above.)								
A state requirement	94	100	93	100	0	100	95	88
Our school system's requirement ...	6	0	7	0	100	0	5	12

TEACHER EVALUATION—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
• Who was responsible for developing the test (or having it developed)? (Percent responding who said "yes" above.)								
The local	7	0	7	0	100	0	5	12
The state	48	100	33	56	0	85	41	36
It is a commercially developed test	44	0	59	44	0	15	55	48
Other	2	0	0	0	0	0	0	4
Which members of the professional staff are evaluated at least once a year—other than by a competency test—on the quality of service they provide? (Percent responding.)								
No one is evaluated periodically	0	0	1	0	0	0	1	0
Only some classroom teachers (e.g., nontenured teachers)	19	12	14	20	22	9	17	35
All classroom teachers	17	24	9	23	14	29	15	5
All building-level professional staff, including principals	11	10	13	10	11	18	9	6
All professional staff	53	54	62	48	53	44	59	54
Are the results of competency tests or other evaluations used for any of the following purposes? (Percent responding. Percents may add to more than 100.)								
a. Not used in this school system								
Competency test scores	67	74	65	68	65	73	74	54
Other evaluations	9	10	12	10	7	9	10	8
b. Reappointment dismissal of teachers								
Competency test scores	2	0	2	3	4	2	2	2
Other evaluations	72	61	77	70	77	74	70	71
c. Selecting teachers for tenure								
Competency test scores	1	0	2	1	2	1	1	1
Other evaluations	51	60	52	51	51	52	51	51
d. Advancement on the salary schedule								
Competency test scores	1	0	0	2	2	1	2	0
Other evaluations	14	24	16	9	14	15	12	15
e. Awarding of merit pay								
Competency test scores	1	2	0	2	2	1	3	0
Other evaluations	4	9	3	2	1	4	6	2

TEACHER EVALUATION—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
f. Selecting teachers for administrative/supervisory positions								
Competency test scores	3	0	7	3	5	1	3	6
Other evaluations	39	36	53	30	43	17	47	56
g. None of these purposes								
Competency test scores	33	32	30	32	37	30	26	46
Other evaluations	16	24	12	13	17	16	15	17

Discrimination

- Three-fifths (61 percent) of the school systems have a WRITTEN plan specifically designed to eliminate discrimination against some groups in some areas.

- Of those 61 percent that have a written plan:

  a. Almost all (94 percent) speak to initial employment and three-fourths (75 percent) to assignments to administrative/supervisory positions. A smaller majority speak to sex role stereotyping in instructional materials (70 percent), and intramural/intermural sports (about 60 percent).

  b. More than four-fifths provide for female employees (87 percent), female students (82 percent) and blacks (81 percent). About seven-tenths (71 percent) provide for Hispanics and some other minorities.

  c. A majority provide for national origin (82 percent), religion (80 percent), physical handicaps (80 percent), age (78 percent), marital status (63 percent), and political beliefs (55 percent). One-third or less provide for sexual preference (34 percent), organizational memberships (33 percent) or life style (26 percent).

DISCRIMINATION

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
Does your school system have a WRITTEN plan that was specifically designed to eliminate discrimination against some groups in some areas? (Percent responding.)								
Yes	61	68	58	53	64	51	62	74
No	39	32	42	47	36	49	38	26
If YES, answer each of the following:								
• Are the followng *areas* SPECIFICALLY provided for in the plan? (Percent responding who said "yes" above. Percents may add to more than 100.)								
Initial employment	94	92	93	93	95	95	95	91
Assignments to administrative and supervisory positions	75	69	79	70	80	62	75	86
Intramural sports	61	71	58	64	48	66	59	60
Intermural sports	63	71	67	65	51	61	64	66
Sex role stereotyping in instructional materials	70	80	74	68	53	70	75	64

DISCRIMINATION—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
• Are the following *groups* SPECIFICALLY provided for in the plan? (Percent responding who said "yes" above. Percent may add to more than 100.)								
Blacks	81	75	90	79	84	66	79	95
Hispanics	71	68	63	[illegible]	80	60	68	85
Some other minorities	71	68	68	61	77	66	69	80
Female employees	87	91	95	85	86	88	88	87
Female students	82	80	90	81	73	81	83	82
• Are the following *characteristics* SPECIFICALLY provided for in the plan? (Percent responding who said "yes" above. Percents may add to more than 100.)								
Age	78	90	74	75	75	86	83	67
Religion	80	90	74	80	80	84	86	70
Political beliefs	55	51	50	59	45	64	56	46
Physical handicaps	80	82	79	75	82	86	76	80
Marital status	63	62	58	65	59	84	59	50
Life style	26	31	13	33	18	33	24	22
Sexual preference	34	36	26	44	29	46	26	34
National origin	82	82	76	77	84	74	88	82
Organizational memberships	33	33	21	35	26	41	32	28

Specialists

The following definitions are used:

"Not available": system does not have such personnel.

"Available only on request": not housed in any schools.

"Limited provision": housed in some schools and available to others on request.

"Fully provided": housed in all regular schools.

- More school systems provide for special music and physical education teachers in the elementary schools than for special art teachers. One-fourth (26 percent) do not provide special art teachers, whereas only 5 and 14 percent, respectively, do not provide special music and special physical education teachers. Furthermore, only 41 percent fully provide special music teachers but 64 and 60 percent, respectively, fully provide special music and special physical education teachers.

- Seven-tenths (69 percent) fully provide special reading teachers in elementary schools, with only 55 percent fully providing them in secondary schools. Only 3 and 8 percent, respectively, do not make special reading teachers available at the elementary and secondary levels.

- School nurses, librarians, and guidance counselors are more likely to be fully provided at the secondary level than at the elementary level. Almost all (96 and 95 percent, respectively) fully provide librarians and guidance counselors in the secondary schools whereas only 65 and 28 percent, respectively, fully provide them in elementary schools. About one-third (30 percent) do not even make guidance counselors available at the elementary level. Two-fifths (42 percent) fully provide a school nurse in secondary school with 37 percent making limited provisions, whereas only 33 percent fully provide a school nurse in elementary schools with 45 percent making limited provisions.

- Most school systems make some provision for a psychologist (94 percent) in both elementary and secondary schools, with only about one-third (32 percent) making full provisions. Only half (about 50 percent) make psychiatrists available and then primarily upon request, with only 10 percent making at least limited provision for a psychiatrist at both levels.

- One-third do not make a social worker available at either the elementary or secondary levels (33 and 35 percent, respectively). About 40 percent make at least limited provision at both levels.

- More than half (57 percent) fully provide a speech therapist at the elementary level and almost half (47 percent) at the secondary level. Only about 1 percent make no provision.

- Teacher aides and unpaid volunteer aides both are more likely to be made available at the elementary level than at the secondary level. About half (52 percent) fully provide teacher aides at the elementary level with another 42 percent making limited provisions. At the secondary level 40 and 38 percent, respectively, make full and limited provisions. Almost seven-tenths (70 percent) make at least limited provision for unpaid volunteer aides at the elementary level, whereas only about half (50 percent) make such provisions at the secondary level.

SPECIALISTS

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
To what extent are each of the following special kinds of SCHOOL PERSONNEL made available to the schools in your school system?								
Not available—system does not have such personnel								
Available only on request—not housed in any schools								
Limited provision—housed in some schools and available to others on request								
Fully provided—housed in all regular schools								
a. School nurse								
Elementary								
Not available	10	3	18	14	6	13	10	8
Available only on request	12	0	18	16	9	16	9	11
Limited provision	45	37	40	46	51	34	45	58
Fully provided	33	61	25	24	35	37	37	23
Secondary								
Not available	10	1	20	14	5	13	10	7
Available only on request	11	1	17	14	10	16	7	11
Limited provision	37	18	34	46	37	36	34	43
Fully provided	42	80	30	26	48	36	50	50
b. Librarian								
Elementary								
Not available	4	3	0	4	9	3	6	3
Available only on request	3	1	0	3	5	3	2	3
Limited provision	29	39	21	25	35	27	31	28
Fully provided	65	58	79	68	52	68	61	66
Secondary								
Not available	0	0	0	0	1	1	0	1
Available only on request	0	0	0	0	0	0	0	0
Limited provision	4	7	3	3	4	7	3	3
Fully provided	96	93	97	98	95	93	97	96
c. Guidance counselor								
Elementary								
Not available	30	33	28	30	32	22	37	29
Available only on request	9	6	4	12	11	11	7	7
Limited provision	34	23	44	33	29	26	31	47
Fully provided	28	38	24	25	28	41	25	17
Secondary								
Not available	1	0	3	1	1	3	1	0
Available only on request	0	0	0	1	1	1	0	0
Limited provision	4	5	1	6	0	8	2	2
Fully provided	95	95	96	92	98	88	97	99

SPECIALISTS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
d. Psychologist								
Elementary								
Not available	6	3	9	5	7	11	5	1
Available only on request	33	23	40	31	44	33	30	38
Limited provision	29	29	23	31	23	29	25	36
Fully provided	32	46	28	33	27	27	40	26
Secondary								
Not available	6	3	9	5	7	12	5	1
Available only on request	34	22	41	30	42	35	30	38
Limited provision	29	30	21	32	25	28	23	38
Fully provided	32	45	30	32	27	25	43	24
e. School social worker								
Elementary								
Not available	33	29	20	26	50	43	31	24
Available only on request	26	25	35	24	23	29	24	24
Limited provision	22	21	26	26	16	16	22	31
Fully provided	19	24	20	25	11	12	23	21
Secondary								
Not available	35	30	21	26	52	46	33	24
Available only on request	27	25	37	26	21	29	25	25
Limited provision	19	19	24	22	14	13	19	27
Fully provided	20	26	18	26	14	12	24	23
f. Psychiatrist								
Elementary								
Not available	51	29	44	53	59	56	45	53
Available only on request	37	57	35	37	35	36	44	29
Limited provision	8	11	14	5	5	5	8	12
Fully provided	4	4	7	5	1	3	3	6
Secondary								
Not available	50	29	44	52	58	54	45	54
Available only on request	38	57	35	38	36	36	45	30
Limited provision	8	11	16	5	5	7	7	11
Fully provided	3	3	5	5	1	2	3	6
g. Speech therapist								
Elementary								
Not available	1	1	5	1	1	3	2	0
Available only on request	11	8	14	10	10	10	8	16
Limited provision	31	30	30	26	40	27	27	41
Fully provided	57	61	51	63	48	61	64	43
Secondary								
Not available	2	1	5	2	1	2	3	1
Available only on request	15	16	17	15	13	17	13	17
Limited provision	35	32	33	33	46	35	30	44
Fully provided	47	51	46	50	40	47	55	38

SPECIALISTS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
h. Teacher aides								
Elementary								
Not available	4	5	8	4	3	7	3	2
Available only on request	2	3	0	1	1	3	1	1
Limited provision	42	37	49	45	40	28	42	61
Fully provided	52	56	43	50	56	63	54	36
Secondary								
Not available	18	15	21	23	15	30	13	11
Available only on request	4	3	3	3	4	6	3	2
Limited provision	40	33	50	40	36	25	41	57
Fully provided	38	49	26	34	45	40	43	30
i. Unpaid volunteer aides								
Elementary								
Not available	18	14	18	20	16	33	13	7
Available only on request	15	21	12	15	13	19	14	11
Limited provision	43	42	41	46	44	29	48	51
Fully provided	25	23	30	19	28	19	25	32
Secondary								
Not available	30	29	26	33	28	53	25	11
Available only on request	16	17	12	19	12	18	17	12
Limited provision	38	41	38	35	43	22	43	48
Fully provided	16	13	25	13	17	7	15	28
j. Special reading teachers								
Elementary								
Not available	3	0	3	3	3	4	3	1
Available only on request	3	0	4	4	3	1	5	2
Limited provision	26	18	26	29	27	14	22	48
Fully provided	69	82	68	65	66	81	70	50
Secondary								
Not available	8	7	12	9	3	18	4	3
Available only on request	3	0	1	4	5	3	4	3
Limited provision	33	26	33	40	29	27	31	45
Fully provided	55	67	54	47	62	52	61	49
k. Special ART teachers in the elementary schools								
Not available	26	8	22	25	35	32	26	20
Available only on request	5	1	7	5	3	3	4	8
Limited provision	28	24	39	26	31	18	24	45
Fully provided	41	68	33	44	31	48	46	27
l. Special MUSIC teachers in the elementary schools								
Not available	5	1	8	6	7	5	7	3
Available only on request	3	1	4	1	4	1	4	4
Limited provision	28	25	34	25	31	15	29	44
Fully provided	64	73	54	68	58	80	61	49

SPECIALISTS—(CONTINUED)

Question	Total	Region				Stratum		
		Northeast	Southeast	Middle	West	300-2,999	3,000-24,999	25,000-or more
m Special **PHYSICAL EDUCATION** teachers in the elementary schools								
Not available	14	4	12	13	18	14	17	12
Available only on request	2	3	1	1	3	1	3	2
Limited provision	24	20	31	22	28	17	22	34
Fully provided	60	74	56	64	51	68	59	53

Ms. Robinson. The suppliers of these group-administered, normal, reference-standardized exams at this point have a relatively free hand in the way these tests are constructed and used based on their appeals to decisionmakers of tests purchased.

We are pleased to note that some reform is currently underway within the testing enterprise itself, and we are somewhat encouraged. However, we do not see that this bill would in any way impose on the legitimate uses to which tests currently have been applied.

In fact, we find that this bill would encourage greater reform and greater scrutiny from within the industry itself.

During this time, the political environment that we might now observe, when a theory and an unproved theory of supply-side economics being applied to every aspect of human endeavor, we would urge this panel to look closely and resist any temptation to apply this economic theory to the yet unsound and unproved definitions of many of the tests that I have indicated are currently in use.

Intelligence tests, aptitude tests, and reading readiness lack clear technical definition within the industry itself. Therefore, our rush to apply objective measure has to be cautious. I find that in this debate we might be faced with a rather common conflict of values, that being a conflict between our need for effective and objective, if you will, measure of our human endeavor and some other less easily measured yet cherished values.

Can we indeed apply high technology industrial model of production to every human activity? And, indeed, if we can, are we willing to pay the cost of that application.

This bill is opposed because it will be financially inconvenient to test developers, and indeed to some test users, but given the choices that we have to make and given the values that we might apply to our decisionmaking process, I would urge greater consideration for the individual, for the productivity and support of our society rather than for financial considerations of industry.

Thank you. I look forward to questions and interaction with the panel.

[The prepared statement of Sharon Robinson follows:]

PREPARED STATEMENT OF SHARON ROBINSON, DIRECTOR OF INSTRUCTION AND PROFESSIONAL DEVELOPMENT, NATIONAL EDUCATION ASSOCIATION

Mr. Chairman, I am Sharon Robinson, Director of Instruction and Professional Development for the National Education Association. I am pleased to have an opportunity to testify in support of HR 1662. We want to compliment Mr. Weiss and the cosponsors of HR 1662 for their commitment to equal educational opportunity. Our members are greatly concerned about the problems of testing. Your willingness to advocate a major change to open up college admissions testing to public scrutiny will assist parents, students, educators, and the public in understanding the issues surrounding testing.

NEA believes that there is an urgent need for the Congress to enact HR 1662. We favor the bill because we believe it will improve access to important information needed by the public and the education profession to scrutinize college aptitude tests. Too many of the nation's youth and their parents are deprived of information about the testing process which affects them directly. Openness in testing or "truth-in-testing" would be a major step in providing students who take these tests with the knowledge of what to expect and allow teachers and parents to have the best possible information available to help students prepare properly.

The NEA believes that all standardized tests administered by school districts must be open to public scrutiny. We believe that the enactment of HR 1662 would be an important step to open up the entire testing process in this nation.

The NEA supports HR 1662 because we believe it will --

- provide information where there has been secrecy;
- serve as the basis for discussion of the serious issues surrounding testing;
- result in improvements in substance and process in assessing

student aptitudes;

- contribute to fairness by providing students with an opportunity to verify their test scores, to compare their responses with those required by the test companies, and to make sure that tests are accurately scored;
- bring about improvements in the tests, and
- lead to the utilization of a variety of viable alternatives.

The 9,000 delegates to the 1981 NEA Representative Assembly adopted four resolutions on testing. They are attached to this statement in Appendix A. In addition to the resolutions the delegates adopted a legislative program which included the following item.

> "Standardized tests should not be used to deny students full access to equal educational opportunity to evaluate students on a single national or state basis. Truth-in-testing legislation should be passed and the Congress should investigate the standardized testing industry."

Those who design, market, and in many cases administer and interpret the tests must be accountable.

They must be accountable because children, the public, and the teachers need protection from tests which much too frequently contain awkward items, ambiguous terminology, and stark inaccuracy. If the tests were not used to make decisions about students so important that their entire life chances may be affected, it might be different. But tests in general are used to rank, group, and track students. They are used to promote and retain students. They are used to grant and deny diplomas. Students are admitted to programs for the gifted on the basis of test scores. They are placed in remedial and retarded classes on the basis of tests. Individual schools, school systems, and total state educational efforts are judged on the basis of tests.

All these things happen in the name of aptitude assessment, more recently renamed developed abilities by some testing agencies. The tests being considered here measure neither aptitude nor general abilities. They are achievement tests pure and simple. One has only to examine items in the verbal part of the SAT. Large proportions of these require vocabulary identification, learnable by almost all students given the opportunity and the time.

In particular, the so-called aptitude tests to which the proposed legislation will apply are used in ways that deny students admission to particular institutions and prevent their pursuing specific career objectives. Such denials frequently reflect an exclusionary posture that is unfair. It is unfair because denials are often based on criteria which represent past socio-economic status and opportunity rather than future potential. The findings of Alan Nairn in the 1980 study of the Scholastic Aptitute Test will illustrate how the test results reflect socio-economic disparities. The study reports that "A Ranking of people by SAT scores remains by and large a ranking by family income."

And it goes on to say that "Although it is advertised as a test of 'scholastic aptitude' and although it is used by colleges to accept and reject applicants ostensibly on the basis of merit, for many students the SAT may be more a reflection of their social class than of their potential for accomplishment inside or beyond the school."

These discriminatory effects of the so-called aptitude tests, particularly on poor and minority students, threaten to rob the society of some of its highest potential leadership, particularly when one considers that test scores do not predict success in later life. In the unique pluralistic society that is America, we need to seek means which will open doors for a broader range of contributions to the body politic from the broadest range of diverse backgrounds which we as Americans cherish. NEA believes this will eventually be best accomplished by substituting for tests a variety of multiple criteria for decision-making about students. But as that is being worked at, we believe the proposals in HR 1662, if adopted, will provide access to the tests themselves and a scrutiny heretofore impossible. Such scrutiny will cause public policy makers and the profession to make haste to employ a broad range of alternatives which, in the long run, will be more accurate, more fair, and more contributing to increasing the pools of talent required in many walks of life in America.

In the broader sense, NEA is concerned that the nation's schools are inundated by by norm referenced, group administered, standardized tests (of which aptitude tests are a part) and that these tests have come to control and narrow the options of local school decision-makers on both curriculum and instructional processes. Local control of education by parents, educators, and elected local officials is being usurped by the products of testing firms, firms which are not accountable to any electorate.

Standardized testing has become pervasive in American education. Students in 88 percent of our schools are required to take standardized achievement tests according to a 1980 study by NEA's Research Division. Only one percent of the school districts reported that they did not administer standardized achievement tests. Standardized intelligence tests are used by 81 percent of the nation's school districts while 95 percent use standardized reading readiness tests. It is clear from NEA's research that there are few schools indeed in this nation which do not administer or require students to take some form of standardized tests.

Congress has passed few education laws which do not require some form of "objective" measures. In almost every instance these "objective" measures are interpreted as standardized tests. In its effort to reconcile the federal budget last month, the House adopted language for "objective measures of educational achievement" which if adopted in the final version would probably lead to an annual nationwide, state administered, locally based standardized test of student progress in basic skills. In ESEA Title I, this approach has not proven to be satisfactory. Local results show a clear improvement on standardized tests, but no satisfactory way has been found to aggregate these scores nationally. Those who have promised to take the government off the backs of the people have adopted an intrusive, narrow, and onerous method of assessing public education.

We are concerned that the direction of education in America is being controlled by standardized tests. Many teachers express the opinion that school districts place too much emphasis upon standardized tests. In a survey of a major school system in Texas, 60 percent of the teachers stated that there was pressure to raise test scores at any cost.

Many of these teachers wrote unsolicited comments upon the survey form, such as, "The pressure of test scores has caused great stress and anxiety." Another said, "We have spent this entire school year teaching for tests. We have reached the point that we are no longer teaching test taking skills but for all practical purposes, we are teaching the test." A third educator said, "I have no time to cover content areas because I am always teaching for the test."

Parents have almost come to accept a computer printout of their child's test scores (ranking him or her against a national sample) as a full measure of the child's education progress and potential. NEA objects to the use of SAT or other standardized test scores as an indicator of scholastic aptitude or as a measure of thequality of American education. We believe that the results of these tests do not accurately reflect either the quality of the schools generally or the learning progress of students specifically.

We remain steadfast in our support of HR 1662 and our opposition to the secrecy and application of norm referenced, group administrered, standardized tests--regardless of whether scores are going up or down. Our considered judgment is that the SAT's do more harm than good.

NEA supports HR 1662 because it will require testing companies to provide information to substantiate claims made for their products. One important example of the need for information is the coaching study done by the Federal Trade Commission (FTC). NEA along with several other education organizations exerted pressure upon the FTC to release data pertaining to the effect of coaching upon SAT and PSAT scores which we were unable to secure from the testing companies themselves. After repeated requests and a Freedom of Information complaint, the FTC released raw data to the NEA. However, it wasn't until this spring that the FTC published its own report on coaching and its effect upon test scores.

The coachability of standardized tests was the subject of an NEA study reported in 1980. A copy of that study has been provided for the record. The study was an independent analysis of an FTC data base created initially to investigate the claim that coaching does improve individual test score performance on the SAT and PSAT.

Based on these analyses, NEA reported a significant difference between the test score performance of coached and noncoached student groups. In all analyses, SAT averages were higher for coached than for noncoached groups. In nearly all analyses, the differences were statistically significant. This being so, one must ask "What of the students whose parents can't afford expensive private coaching schools? What kind of fairness doctrine does this provide them?" Parental income was the one characteristic external to SAT and PSAT scores which appeared as a significantly discriminating variable. For example, the greatest average gain reported for total scores was found for students in 1976 who were coached between the PSAT and the first time they took the SAT and whose average family income was $29,000. For this group an average gain of 143 points was found. This confirms the Nairn findings I cited earlier.

Mr. Chairman, the NEA endorses openness in all standardized testing. We believe that standardized testing has become accepted as an almost infallible measure of education progress by the media, by parents, and by students who are caught in the testing "secrecy trap." The example alone of their coachability discredits such infallibility.

NEA supports HR 1662 because we believe that its enactment will provide a first step to opening up all standardized tests to public scrutiny. It will provide information where there has been secrecy. It will provide accountability where there has been arrogance. It will provide a basis for productive dialogue on college aptitude tests based upon relevant information and data. Above all, it will contribute to the fairness of these tests because it will provide all students with more equal opportunity in preparing for and submitting to the tests. It will provide elected officials, professional educators, parents, and students with the information necessary to judge whether or not standardized testing as presently practiced should determine the future educational and career opportunities for millions. And we are prepared to abide by these judgments, because we have faith that reasonable people will conclude, as the research cited earlier has concluded, that the tests are more a measure of social class than student potential.

APPENDIX A

E-20. Accountability and Assessment

The National Education Association recognizes that the term *accountability*, as applied to public education, is subject to varied interpretations. The Association maintains that educational excellence for each child is the objective of the education system. The Association believes that classroom teachers can be accountable only to the degree that other parties who share this responsibility--legislators, other government officials, school boards, administrators, parents, students, and taxpayers--are also held accountable.

The Association believes that there should be no single or statewide accountability system. It will resist any attempt to transform assessment results into a national or state testing program that would seek to measure all students, teachers, or school systems by a single standard and thereby impose upon them a single program, rather than providing opportunities for multiple programs and objectives. The Association believes that specific behavioral objectives should not be used as course objectives, nor as a basis for determining accountability.

The Association opposes the unquestioned pursuit of behavioral objectives and insists upon --

a. A critical examination of the effects of use of behavioral objectives

b. The identification of areas where behavioral objectives are useful but not harmful

c. The acceptance of alternative statements of objectives

d. Recognition that evidence of learning is not always available upon request or demand

e. Recognition that performance criteria are not always uniformly or universally applicable to a given population. (71,76)

H-10. Testing

The National Education Association recognizes that testing of students, preschool through job entry, may be appropriate for such purposes as --

a. Identifying learning needs

b. Recommending instructional activities

c. Describing student progress.

The Association opposes the use of tests that deny students full access to equal educational opportunities or that are used to evaluate teachers.

The Association believes that standardized tests should not be administered when they are --

a. Potentially damaging to a student's self-concept

b. Biased

c. Used as the only criterion for student placement

d. Invalid, unreliable, or out-of-date

e. Used as a basis for the allocation of federal, state, or local funds

f. Used by testing companies or publishers to promote their own financial interests at the expense of sound educational uses

g. Used to compare individual schools

h. Used in an exploitive manner by the media

i. Used as the sole criterion for graduation or promotion

j. Inappropriate for the use intended

k. Used as a criterion for the development of a state system of classification of schools and/or school systems. (78,80)

H-11. Criterion-Referenced Tests

The National Education Association believes that criterion-referenced tests are a viable alternative to standardized norm-referenced tests. Such tests should be designed to describe student performance based on carefully developed curriculum. It is inappropriate to administer criterion-referenced tests that do not specifically measure instructional content.

Staff, time, instructional materials, and other resources should be provided to assist students who experience difficulty achieving the desired criteria reflected by tests. (80)

H-12. Truth in Testing

The National Education Association believes that intelligence, aptitude, and achievement tests have historically been used to differentiate rather than to measure performance and have, therefore, prevented equal educational opportunities for all students, particularly minorities, lower socioeconomic groups, and women. Contemporary research on the structure of the intellect identifies multiple and varied mental operations and advances the significant premise that these operations can be taught, that intelligence is dynamic rather than fixed.

The Association further believes that the truth-in-testing movement is an important step for bringing about long-needed test reform. Therefore, it urges all state affiliates to strive for passage of truth-in-testing legislation that includes a provision for each individual test taker to receive a copy of all test questions, scores, and rationale for correct answers. (80)

Mr. Weiss. Thank you very much, Dr. Robinson.

First of all, Dr. Casteen, let me express my enthusiastic support of the effort that you have described.

The only question I have is that I thought that you had mentioned in your statement that you had a list of the proposed authors or researchers, but I don't find that.

Mr. Casteen. I may have spoken imprecisely. We have described the kinds of people with whom we are negotiating. Our concern of releasing a listing of persons with whom we are talking before we have a commitment may drive away competent persons whom we would like to recruit.

Mr. Weiss. The only potential problem is who is selected to do those studies.

Mr. Casteen. In the prepared statement I explained the areas of possible mismanagement that we see in the project. We see this as a volatile area and we think parties on both sides of this discussion will have ample notice to try to win a debate, if this becomes a debate.

We don't want to go that way. We want to try to engage academically and professionally competent persons who can work with the guidance of a steering committee, the kind of committee in which we would hope at some point to involve persons like you, to see to it that what is produced is a thoughtful, dispassionate document that might indeed serve as a basic public document. The pitfalls are there.

Mr. Weiss. I think it is a very welcome approach.

One of the aspects of this whole discussion about this legislation and its predecessor that I commented on previously is the disconcerting sight of otherwise independent, intelligent professional people taking marching orders as to what their position ought to be on a specific issue.

Now, my own prejudices lead me to believe that they were all primarily on the part of those opposed to this legislation, and I know that whenever the College Entrance Examination Board senses that there was a hearing coming on or legislation being introduced, letters and telegrams would be sent to college and university presidents telling them what their positions ought to be and, almost uniformly, those college presidents—without any independent valuation of what the legislation was—would feed back to us, Members of Congress, precisely the arguments that ETS or the bureaucracy of CEEB had decided upon.

That is a problem and I hope your study can lend some light and objectivity to this whole question.

Mr. Casteen. I think the object of it is crucial. I don't really know enough to argue either way, but it may be useful to observe that the other side has also tried to recruit probably the same population to argue on the opposite side of the question.

It is a fact that we in education who have some kind of ongoing stake in the quality of what happens to children in schools are fearful of changes that may have the effect, and I have to emphasize "may," that may have the effect of taking away the indicators that we have been able to use to try to judge where we are and where we need to go, and it is the uncertainty about the effects of the legislation and the uncertainties of separation of what I see as an important mandate.

My association, NACAC, honestly needs guidance of the kind we can provide in the sort of report I am describing.

The same report may be useful to you, and it is worthless to us if it is biased at the outset. It has to make some reasoned judgments as to how we as practitioners who must make decisions can best deal with the interests of our institutions, society, the student, and so on.

Mr. WEISS. I think that Dr. Sjogren had indicated that he had read the prior record of hearings, and I think that again, if your people do that, you will find, not answers necessarily, but you will find the questions posed by both sides to this debate and discussion, certainly as a guide to what some of those questions ought to be that need to be answered at this point.

Let me just ask as a start-off one question which has been very, very critical to this whole discussion; there is a factual dispute on this, and it seems to me that there ought not to be.

Sometimes I don't know whether it is factual or semantic. We have heard on at least two occasions in the course of this panel, I think Dr. Sjogren, you know of no one, or no admissions officer who uses the test scores solely as the determinant for admission. That statement is made repeatedly.

At the same time, and I am going to put it by way of a statement but I would like it to be interpreted as a question for a response, at the same time I know as a matter of fact on the basis of statements in the college catalogs, those that summarize the various colleges and universities and indicate what their areas of specialization are and what is needed to be for likely favorable consideration, that there are cutoff scores indicated, and it seems to me it is only half an accurate statement to say that the test scores are not used solely to determine admission, when in point of fact the test scores are used solely for denial of admission.

I would like all of you to respond to that; is there or is there not the utilization of test scores for denial of admission to colleges and universities?

Mr. SJOGREN. I too have read of your and others' concerns on that issue and looked at catalogs, and there may be a misunderstanding. I am again speaking from the undergraduate point of view.

At the undergraduate level I, too, have observed in some catalogs that there is a minimum test score that will allow entry into a university, but usually that statement is a secondary consideration if the student does not meet the established standards on the basis of a high school record.

Mr. WEISS. It does not say that this will give you unfavorable consideration for admission. What it tells the perspective applicant is that if you score below this level, do not bother applying to that school because you will not be considered for admission.

Mr. SJOGREN. I would be very disappointed if that did happen. It would certainly run counter to the practices constantly discussed in annual meetings and through professional literature.

There will be scores below which students will not be admitted if those students have not met the other standards.

A very prominent western university requires students be in the top 12½ percent of their high school class. If they do not qualify, there are other ways that they might qualify, and one of those other ways is through cumulative scores on a series of tests, and virtually all of those institutions beyond that will have special considerations given to students from minority groups, older students, low income populations, rural populations, so these statements, while they are in the catalog, sometimes are not the ones that are actually practiced.

I would be disappointed if an institution at the freshman level did require a single test score for admission.

Mr. WEISS. Dr. Casteen?

Mr. CASTEEN. Just a series of not well-ordered thoughts on the question you asked.

As Mr. Sjogren indicates, the test user advice that we receive from the test makers and sponsors with whom we deal always tell us that that use of a test is inappropriate, so begin with that.

Having filled out the questionnaires on which many of those standards are based, I should observe that often the way they ask the question predicts the answer.

For example, if the answer says, what is your cutoff score, and if I mark through cutoff score and put average or mean or ranking of students commonly admitted, that, in fact, indicates the extent to which we do not do that, the publication indefinitely goes back to the phrase "cutoff" score, and my effort to give a valid picture of what really happens has been abbreviated.

To the extent that colleges may misuse test information, and I agree with my colleague, anyone who looks at the situation, that there are some instances of that, H.R. 1662 does not correct the problems. It has to do with the test sponsors.

If, for example, the cutoff scores are used in admissions, and I can make that from "if" to "since" cutoff scores are used in awarding financial aid, a condition about which I am as alarmed as my colleague to my left, that use clearly violates the directions given by the sponsoring organizations and professional agencies.

The legislation does not address that, and so the jeopardy in which students may find themselves when bad practice prevails would survive this kind of legislation.

Mr. WEISS. The legislation does not cure it because the legislation is not intended to mandate any course of action as far as admissions practices are concerned or utilization of the test.

All that the legislation seeks to do is to mandate disclosure, so that the student, the academic community, the world at large knows what those tests are about. Implicit in that is the concept that if, in fact, all the information is out there, the cures will follow, because of the normal dialog and intercourse that will take place within the community affected.

In my judgment it is not the role of Congress to tell you how to use tests or any other criteria as far as admission practices are concerned or awarding of financial grants or so on.

Mr. CASTEEN. I think you are right, and correcting the problem is the business of education. I would simply observe that in this area, the use of cutoff scores and so on, evidence to the contrary or advice to the contrary has been available for years, and to the extent that it goes on, we probably have evidence of the people who make decisions don't always follow the best advice given to them.

Mr. **Weiss**. Perhaps one of the areas where the study that you spoke about can be helpful is in fact focusing some new attention on this problem and determining factually whether in fact the charge is valid and to what extent, if it is valid.

Anybody else want to comment on that aspect of it?

Let me touch, if I may, Mr. Shields, on your discussion of the utilization of the test pursuant to the New York law in the preparation courses.

What kind of response have you received, within the community, in terms of participation in those preparation courses compared to what existed previously? How are they set up? What kind of fee structure if any exists?

Mr. **Shields**. Sir, there are no questions for our program, nor are there any fees nor any classes that we might have connected with the program.

Students come to us basically, the majority, wanting to go on to college. They are sophomores, juniors, seniors in high school, and they come in and say, "I want to go to college. What do I need to do to get there?"

In order to help prepare them to do so we have established seminars that take place before each one of the testing dates, 3 weeks before each one of the testing dates and any student is eligible to those free of charge.

The purpose of them again is not to teach those students how to read better or write better, et cetera, but rather simply to get them familiar enough with the tests so they are walking in, and they know what they are getting themselves into, because with the type of student I work with, the problem we face most often, kids score 200 flat on the verbal, which means a zero basically when their competency is much higher than that, and so it is a problem with taking the test.

Since we have been able to use actual copies of the SAT's, attendance at these seminars has doubled and for the October test date, which is the largest test date for most students of my group, it has tripled, in other words, the response has been much greater, and I also have in my hand letters from another talent search agency, Bronx Community College in New York City, likewise indicating that since they have been able to use these seminars, there has been an increase in attendance and a more positive reaction on the part of the students.

Mr. **Weiss**. We have received testimony in the course of these hearings, testimony on the so-called preparation, coaching schools.

We had some testimony yesterday to the effect that in fact if those tests are utilized by way of preparation and coaching, the people who are going to benefit from it are going to be the wealthier students, because they say now the coaching schools which charge anywhere up to $300 or more for those courses will have more material available and it will make it easier for those who can afford those courses, and I would like your comment on that.

Yours is the first testimony we have had in the course of these 2 years of hearings as to community-based preparation or coaching schools and I wonder if you have any comments about that argument, that really the benefit accrues to the wealthy rather than to the disadvantaged.

Mr. Shields. There is a question of magnitude, on-the-margin benefit.

Mr. Kaplan, whom I know, does an awfully good job with his program; he has some very, very good materials, and he has been working for a number of years to develop them. His getting a hand on a real copy of the SAT is going to help him, but it is not going to help him nearly as much as it will help a program like mine, whereas before we got our hands on the SAT all we could do was go by the prep guide.

We have taken a jump here in our ability to work with those students whereas Mr. Kaplan will take a small step.

Mr. Weiss. A number of you, Dr. Schafer, among you, have indicated an attitude toward disclosure, which seems to be more conservative than that which is the current practice of the Educational Testing Service.

Now, would you care to comment on that?

Do you oppose the College Entrance Examination Board, the ETS policy of disclosing five out of seven major administrations?

Mr. Schafer. I think in general the Educational Testing Service has for whatever reason chosen to disclose, and I think perhaps some of this may be in response to criticism as well as the LaValle law in New York State.

With respect to disclosure, my comment has to do more with the balances of the advantages and disadvantages. There is a little doubt that better testing enhances accuracy, evidences that disclosure does enhance accuracy.

The gain is minor and the cost is major. Balanced against greater costs, the argument for disclosure becomes far weaker than the argument against it.

Mr. Weiss. You would think ETS ought to reverse the policy and go back to where it was previously?

Mr. Schafer. As I understand it, that would disadvantage students in New York State who would then not be allowed to take the test. Whether or not that is an advantage, I care not to comment on.

Mr. Weiss. Dr. Hanford, who is the president of the College Entrance Examination Board, testified yesterday that the disclosure policies in effect now have been an actual outgrowth of the direction in which they have been heading in the last 25 years or so.

You don't consider the current policy to be a national outgrowth of which the College Board has undertaken over the last quarter of a century?

Mr. Schafer. I do think it is a step a bit too far.

Mr. Casteen. I should make clear that the decision to disclose on the limited basis currently in effect was a College Board decision undertaken under the terms of the contract between the College Board and the ETS, and I should also say as a trustee of the College Board, that I voted for that proposal with a good bit of uncertainty as to its ultimate effect.

We believe that the public interest called for the step in the direction of the disclosure. We went as far as we could within our present financial resources and other educational resources in the direction of disclosure, but frankly it is a risk.

Our experience tells us that there may be, probably over a fairly short run, some deterioration in the quality of the tests that we were able to provide to our client colleges, and so on, and we enter into this also with some sense that this risk is a fairly large one because the extent to which the SAT score can be used as an ongoing gage of educational progress.

We need to be more certain than we are presently able to be that complete disclosure would protect the year-to-year comparability of test scores across time as a way of seeing to it that we can make the kind of diachronic analysis that was made in 1975 and 1976.

Mr. WEISS. Ms. Robinson?

Ms. ROBINSON. The comment that I would offer at this time has nothing to do with the degree of ambiguity and the degree of uncertainty around the use of these test scores and the various procedures in which the test might become involved.

I must note that in the admissions process a number of caveats seem to prevail in terms of how the score might be used, in terms of how the score is to be regarded, and as it relates to disclosure around an instrument that has some uncertainty and insecurity in terms of accepting the instrument, that score, for what it might tell you, causes me to seriously question the test agenda overall.

It was stated during the course of the panel presentations that perhaps there are some more significant underlying questions that must be considered, and I guess I would have to echo that concern in that have we indeed decided the amount of standardization that we want to apply to the overall education through this country?

Have we agreed on the content or the curriculum that these instruments will seek to measure in terms of student progress, student learning?

Have we indeed decided that we can make a distinction between student aptitude and student achievement?

These are very significant questions, underlying questions, and I think in this age of high technology, the ethical question of where we apply what is technologically possible to these decisions that we must make as members of a society is going to be of critical importance to the panel, to the Congress and certainly to educators.

Thank you.

Mr. WEISS. Mr. Erlenborn?

Mr. ERLENBORN. Thank you, Mr. Chairman.

While listening to this testimony I was reminded of an old story, and I would like to preface my questions or comment by retelling it.

Mr. WEISS. Is it clean?

Mr. ERLENBORN. Very clean. It is somewhat dated in its reference to the cost of election campaigns; but the story goes: The young man went to the head of a great university and said: "Sir, I desire to be a captain of industry, and I would like to run one of the great railroads in this country. What course of study should I follow and how much will it cost me?"

The university president said: "Well, such a course of study would take many, many years and cost thousands and thousands of dollars. I would suggest, young man, if you would just devote 6 months of your time and much less money, you could run for Congress and, when elected, you would feel competent to run all the railroads in the United States."

I think that has some bearing on why we are here today. It leads me to ask a question of all the panelists:

Is there a problem that can only be solved or best be solved by legislation and, if so, is the Congress of the United States best equipped to provide that legislation?

Ms. ROBINSON. I will begin. Yes, I think there is a problem that can be addressed by this legislation.

The problem has to do with consumers, with the consumers' interest in the whole endeavor of testing, measurement, evaluation.

I guess the interest of the Federal role here in terms of deciding on proprietary and the appropriate degree of Federal intrusion causes me to suggest that we have a set of principles and values which represent part of the foundation of our society, a part of our foundation of this democracy.

I think it is entirely within the purview of the Federal Government to articulate legislation and regulation which encourages, which helps to interpret an understanding of those values, and indeed which helps us to apply those values in every aspect of our human endeavors.

I see this bill as being in concert with that overall responsibility of the Federal Government and, therefore, the NEA has supported truth in testing legislation, if you will, certainly this bill, H.R. 1662.

Mr. ERLENBORN. Throughout the hearings there has not been a very clear line drawn between the complaints about the openness of the various tests, and the utilization of the tests.

Much of the testimony has been directed to the utilization of the tests and whether they are the sole criterion for admission. Based upon your answer, can I assume, if there are problems in the utilization of these tests, that you would again see the Federal Government legislating the best solution?

In other words, do you think the Federal Government should get involved in the practices of our educational institutions as to admissions?

Ms. ROBINSON. Are you speaking in terms of the colleges?

Mr. ERLENBORN. When and how do you use the tests, how much value you place on them?

Ms. ROBINSON. Certainly not, but indeed in Federal legislation, in the numbers of areas that relate to education, there is an indication that you will depend on some standardized measure of student performance, then indeed you also have a corollary responsibility to see that that measure has integrity and since that same measure which is often required in several Federal programs is like measures required in other aspects of education, we are talking about similar kinds of instruments.

Therefore, achievement tests, which are some measure of student performance, is often required in Federal programs, that is immediately interpreted in most instances as the need for some standardized tests, nationally normed referenced standardized test groups administered and this subjects the entire industry to rules

293

and procedures which suggest the Federal standard of integrity and quality.

Mr. **Erlenborn**. I would like to have your comment, but first let me say that I am a bit surprised to hear a representative of the NEA suggesting a national testing standard.

I thought you would have a little different approach. I might suggest, if we begin to do this with students on national testing standards, maybe performance standards for teachers will follow, a national performance standard which might supplant the tenure that is so highly criticized.

Mr. **Sjogren**. I much prefer to look at this as a critical issue, which has within this issue several problems but also a good many of strengths that we must preserve.

I feel that the great media attention that has been brought to this issue in the past 2 or 3 years, the work of this committee and State committees and interest groups, has caused those of us in the institutions and agencies and associations to really take a good hard look at ourselves, and start evaluating those problems and developing strategies to respond to those and remove those problems.

I have great faith in our system to be able to respond to that and do it effectively without influences from the Federal Government, and I would plead with the committee to allow us to try this, because great steps have been taken and accomplished in the past 2 or 3 years due to a great extent to the efforts and attention that you have already brought forward to this, and I would like more time to allow us to see if we can't correct this ourselves.

I think we can.

Mr. **Erlenborn**. May I ask your opinion as to the value of diversity in these tests. Should we move toward some national, congressionally endorsed testing standards and have then, in effect, one test rather than a diversity of tests; would that be a good move or bad?

Mr. **Sjogren**. I think it would be a bad move. At my own institution we do not rely on one test but several different ones.

Mr. **Erlenborn**. What about the progress of the testing standards and so forth? Is there a level that has been achieved today? We don't want to get frozen in place either, do we?

Mr. **Sjogren**. That is correct.

Mr. **Priddy**. First you ask, is there a problem. I spoke earlier before you were able to join us and in my opinion I indicated clearly there is not.

You asked what contributions does this bill, H.R. 1662, make. I have difficulty seeing those contributions. If you take section 3, a number of things are laid out that would be made available. I think those things are already available.

Take, for example, income; there is a subsection on reporting and income as it relates to achievement on these tests or performance. I think we already know the correlation there. It is very clear.

An educator is an educator is an educator; my response is, so what? What are we going to do about it now that we know he is having some difficulty?

Section 3 is a clear infringement on the private sector and totally unnecessary for the Secretary of Education to have another stack of reports to read each week, and those are already available in the general press.

Section 5, disclosure essentially is what it is about. The business of developing good tests is very complex and to throw all of that away every year is a waste. The costs that would be associated in terms of time is unreasonable, but the cost in terms of money to test takers will be exaggerated beyond reasonable proportions.

I am at a public school system, and we feel it is our business to be responsible to our communities and to our boards for instruction and so on. Testing is only one of the means that we use to be responsible to them. We use the results of group and individual tests to make decisions about curriculum, instruction, and so on.

We do not use them as a sole indicator, indications where curriculum needs to be revised. It is only one of the means that are used, and we are not having any trouble at this point using what is available to us to make those decisions.

Mr. **SCHAFER**. In the testimony that I gave I mentioned a document that the association has produced. In this, our association takes the position that we do not have an urgent problem, and I should also note that with respect to measurement-oriented professional associations, I know of several who have taken a position which is opposed to legislation with respect to testing, specifically disclosure, and I know of none which has taken a position in favor of it.

There are several things in our statement which speak to, at least a beginning, with respect to look at the problem of test use.

Mr. **SHIELDS**. You asked the question whether or not there was a problem. There are a number of problems in education, and just as obviously most of them should not be legislated by the Federal Government. Education is a State function.

Mr. **ERLENBORN**. I like to be reminded of that from time to time.

Mr. **SHIELDS**. Tenth amendment. However, this bill does not try to legislate how admissions officers should use tests, nor should it, but if tests are going to be used and if there are going to be such things as cutoff scores and if tests are going to be such an integral process of the admission and financial aid process, more important then we have to ask ourselves, is there anything that any of us can do which might help those students who have to go through this process?

You cannot tell an elementary and secondary school how to teach their children. You can't tell admissions officers how to use those tests, but I do think it is possible for you to be able to provide more information to the students who have to take those tests, especially those students who would not receive it otherwise.

Yes, there are things out on the market concerning tests. They cost money. My students come about 70 percent from welfare families, and they buy no books, certainly not commercial testing guides. One of the nice things that ETS has done is that for those students who receive fee waivers to take the examination they can receive their test scores back for free.

Basically, what I am saying is that the children that perform most poorly on those examinations could benefit most from having more information about them in their hands and requiring them to buy things, requiring them to send in special forms to get that, it is not going to get those kids that information most quickly, and I think this legislation will.

Mr. **Erlenborn**. Thank you, Mr. Chairman.

I have no further questions.

Mr. **Weiss**. I think that perhaps in the course of the last couple of responses we have gotten some clarification of the thrust which the legislation has.

My distinguished friend and colleague is a marvelous devil's advocate, and sometimes includes in his questions premises which you never intended, and he knows that in fact they are not actually premises but he does it to provoke responses, and I think that is good.

Dr. Robinson, perhaps you ought to amplify the comment that you made because, as it was left as a response to Mr. Erlenborn's question, it put you and NEA in exactly the opposition position of where you are.

Ms. **Robinson**. Certainly in the interest of consumerism, and if we would view the bill in that manner, I would just recall that as consumers have become better informed about the various products that we require, the quality of those products coming under closer scrutiny has helped to encourage better quality indeed.

I suspect that would also be the case with standardized tests, and this bill would encourage better quality of tests.

I do not fear it and I have great faith in the technological skills of those who help to develop tests, and I feel that disclosure will simply encourage all of us to become much better experts in all the endeavors that we undertake in this regard, so while the bill does not indeed, it does not even attempt to establish a Federal standard for the tests themselves, it does establish a value in terms of protecting the consumer, the interest of those who take and use test information.

Mr. **Weiss**. In the course of your prepared testimony, on page 5, you have a statement in which you say:

> Congress has passed few education laws which do not require some form of "objective" measures. In almost every instance these "objective" measures are interpreted as standardized tests. In its effort to reconcile the Federal budget last month, the House adopted language for "objective measures of educational achievement" which, if adopted in the final version, would probably lead to an annual nationwide, State-administered, locally based standardized test of student progress in basic skills.

Are you saying in the course of that statement that in fact it is not those who are asking for disclosure legislation but those who seem to be opposed to it who have now gone the full route and are in fact attempting by the legislative provisions already adopted to put us on a national standardized test basis?

Ms. **Robinson**. We wonder if perhaps that is not indeed beginning to occur, and we have to ask ourselves then, when did we determine the level of standardization that we are ready to accept through mandates such as this?

I think that we have a need for objective measure of all of our work, and we have procedures. We know we have the technology necessary for educators to determine objectively a measure of student progress without such heavy reliance on nationally normed reference group administered tests.

They are, however, convenient and the economy of scale becomes very appealing, so we often give in to that value rather than take the time to question other underlying considerations, such as making a careful examination of the match between the test content and the local curriculum or how are we going to accommodate differences determined by the local community or by the State as it relates to what you will teach and what those various tests will measure.

Certainly, I know that school administrators and educators, teachers themselves all across this country spend a lot of time in trying to make careful selections of those instruments so that they are indeed applying the most appropriate one for their needs and for their students, but the degree, that same amount of time, that same amount of money might be better invested in a kind of assessment procedure which is developed locally, which teachers have greater control over in terms of applying when they have the need to apply, which can be applied whenever they choose, and can provide the kind of information they need on an ongoing basis to truly assist student need and help to measure student progress.

This does indeed suggest that we are going to recognize perhaps the validity and a national measurement scheme that we have not thoroughly examined. The national assessment project notwithstanding, we know that a number of accommodations are made so we are not reporting inaccurately to the public.

This I am not sure how we will implement.

Mr. ERLENBORN. Mr. Chairman, might I say I accept the gentle chiding of the chairman as to the premises of my questions, and I think I owe you an explanation.

I guess I just failed to articulate part of my thought process.

After 25 years as a legislator, I have learned something about what happens to legislation. I have observed it over the course of the years.

It is my opinion, once a legislative body establishes the principle that it has the right to regulate an industry it seldom is reversed, nor does it stand in place. It grows and grows and grows, and I guess I just didn't articulate that too well.

Mr. WEISS. Well, I am glad that you clarified your position, Mr. Erlenborn.

I have one final question.

Dr. Sjogren, in the course of your statement, I am not sure if I caught the full impact of what you meant when you said that companies, I assume testing companies, should be punished if they misuse the test in some way.

Clarify what you meant by that.

Mr. SJOGREN. I am sorry for that ambiguity. What was intended there was to state that testing companies who do not act responsibly such as, I believe, ETS is now acting, the American College Test, the College Board, those are three of the groups with which I am most acquainted, I feel they are acting very responsibly and should be licensed to continue on in doing the good things that we are doing.

However, there are testing companies that fail to act responsibly and fail to meet most of the conditions specified in your proposal, and those companies should be made to, should be punished. That was my intent.

Mr. **Weiss**. Well, you know and I have been sometimes taken to task for appearing to be too critical about ETS.

The American College Testing group has in fact to this moment to the best of my knowledge refused disclosure of any kind. I am not sure you would not include them as one of those who should be subject to some punishment.

Mr. **Sjogren**. The item of disclosure was exempted from my point.

I still have very mixed feelings about disclosure but I am talking about the other kinds of information that is provided to the consumer prior to the testing.

Mr. **Weiss**. Before we conclude our session this morning, if there are any of you who have some additional comment or comments that you care to make just to total out your presentation, so that we don't have you hanging, feel free at this point to do so.

Mr. **Casteen**. This question of the Federal interest in testing and especially college admissions testing is one that has interested me for a couple of years.

I didn't answer Congressman Erlenborn's question earlier because I am really not sure where I would come down on that one on the end concerning some of the things Ms. Robinson said on behalf of NEA.

On page 2 of the resolutions appended to the NEA presentation, it is said that the NEA opposes the use of tests that deny students full access to equal educational opportunities. The extent to which that happens needs to be examined. That is the kind of matter we intend to look into seriously, but that is one point.

Second, as to evaluation of teachers, that is a State function. You asked about State interest in this type of topic, and then the examples of tests that should not be administered, and also Ms. Robinson's expansion on this matter of locally based assessment together I think concern me more than I realized.

The first example given is tests that are potentially damaging to a student's self-concept. Because we must administer 9,000 or 10,000 rejection letters to college, we deal with those people regularly because they have been told no, and it is worth exploring the possibility of the student who does not read well enough, who reads poorly, he ought to know that at a rather early age than local assessment guarantees deliver that information.

The trend toward grade inflation is essentially a locally based form of self-perpetuation of reward. That national measure that might provide a notion as to whether students do read well enough, that that may have some societal value that needs to be weighed.

Students who have not benefited adequately from schools, those who come from ethnic groups, those students deserve early notification, early information of a kind that I fear is not available through most of the local assessment programs that I know about, so I am concerned that the Federal interest in this question may well have to do with protecting the quality of experience of the children who go through schools or in some other way protecting the teachers or the other interest groups.

I am not sure that the Federal interest is defined in this particular set of objections to certain kinds of testing that Ms. Robinson has presented. In the end, the attention focused by your interest in this topic on the whole question of testing has been healthy for education, but I am apprehensive that our best as a nation is served by legislation that may serve to diminish the amount of nationally referenced measurement information that is available to us as we try to assess our progress in schools.

Mr. WEISS. Your statement leads me to ask this question.

Are you at all bothered by the confusion that may well be created in the minds of students as well as their parents by the use of the term "aptitude" when in fact it is achievement that is meant by that word in Scholastic Aptitude Tests; does that bother you at all?

Mr. CASTEEN. That may be a loaded question. I recall that in 1979 I suggested to your committee that what we are looking at is a certain set of skills that can be used to learn, reading, and so on, and a certain set of bodies of knowledge that can be demonstrated on certain kinds of tests.

Personally, I do not find the word "aptitude" especially useful. It is as ephemeral for me as is the word "intelligence." I know what skills are, certain bodies of knowledge are, and I am concerned about the convertibility of those phenomena to educational progress or achievement.

I suspect that the word itself means what one wants it to mean, aptitude, and that we may in the future want to take a good look at that, but I doubt that there is a Federal interest in that word.

Mr. WEISS. Well, the only reason I asked the question is because you confronted the issue raised about people's self-worth, and we have had testimony in the course of these hearings over the last couple of years about how people having taken those tests which they are led to believe because they are told that no amount of coaching or preparation will help them, that, in fact, it is some vague, inherent aptitude for academic pursuits that is at stake and when they don't do well, it is not that they have not been taught well, that they have not studied hard enough, it is that they themselves are not good enough inherently.

That is the problem and, again, you are right. I don't think that I want to mandate by law that you can't use the word "aptitude" but I would think that the educational establishment, both test makers as well as test givers, would want to think real hard as to whether they are not doing themselves as well as the test takers a grave injustice by allowing others to misinterpret what is meant by that word.

Mr. CASTEEN. Good point, I think.

Mr. WEISS. If there is anyone else that wants to comment on anything at all, now is the time.

If not, I very much appreciate your participation this morning.

These hearings will undoubtedly continue and hopefully sometime in the near future, not the day after tomorrow, we will have some further legislative progress to report and I look forward to the reports about the study which you spoke about.

Thank you all very, very much.

The subcommittee stands adjourned.

[Whereupon, at 12:35 p.m., the joint hearing of the Subcommittee on Elementary, Secondary, and Vocational Education and the Subcommittee on Postsecondary Education of the Committee on Education and Labor adjourned.]

[Additional materials submitted for the record follow:]

EDUCATIONAL TESTING SERVICE ETS PRINCETON, N.J. 08541

609-921-9000
CABLE-EDUCTESTSVC

August 6, 1981

The Honorable
Paul Simon, Chairman
Subcommittee on Postsecondary Education
320 Cannon House Office Building
Washington, D. C. 20515

Dear Chairman Simon:

I am writing to you today to request the inclusion of additional material into the official record of the hearings held on July 21 and 22 on H.R. 1662, the Educational Testing Act of 1981. As you know, I was present as a technical adviser at the hearing on Wednesday, July 22, accompanying Dr. Michael Priddy of the Guilford County, North Carolina school system, a witness on the "Experiences of Test-Takers" panel. Inasmuch as the hearing record is being held open for ten days following the hearings, I would like to take this opportunity to supplement the record with the following:

(1) A correction to the testimony of Michael Galligan, a witness on the "Experiences of Test-Takers" panel on Wednesday, July 22. Mr. Galligan, a student who found a flaw in a question on the SAT, indicated in his testimony that because of his finding the error, the grades of almost 20,000 students who had chosen the alternative answer were raised ten to twenty points. He further stated:

> "Unfortunately, students outside of New York, where the "truth in testing" law is not in effect, have not had their scores adjusted. Are not the students in those other states just as correct as the students in New York?"

This statement is grossly misleading because the particular question challenged by Mr. Galligan was not included on test forms used in any other states, and therefore no other students' scores should or could have been changed.

(2) A response to statements made by Michael Galligan during the Question-and-Answer period following his testimony. Rep. Weiss asked

both the student witnesses to comment upon their experience as to students' attitudes about tests, their impressions of whether sufficient information is provided about them, whether students have an opportunity to study for them, etc. Mr. Galligan indicated that prior to taking the SAT, he "had very little idea as to what specifically the format could be..." He explained that he had not taken the PSAT so that, "I had no idea in that respect, nor did I receive information from my school as to what the format was."

Mr. Galligan then reported that students who attended the Jackie Robinson Foundation Youth Convention at Columbia University a few months ago also "did not have the information, the appropriate information, as to what the test would actually test, as to what the format would be and what the necessary levels of math and verbal skills had to be attained so that one could actually do the exercises of the test and display his knowledge in that manner to make the test valid (sic)."

3 We at ETS were astonished at hearing the above statements by Mr. Galligan because of our extensive distribution to high schools and to individual student requestors of material describing in detail the nature of the SAT, the skills and abilities measured, the format of the various question types, suggestions as to how to approach answering the questions, as well as hints for using time efficiently and for guessing. This information is included in the booklet entitled, Taking the SAT (Attachment A) almost three million of which are sent each year to high school counselors across the country for free distribution to students along with test registration materials. Included in the booklet is a complete sample test, with a sample answer sheet and the list of correct answers, as well as an explanation of how to score the test. The memorandum to Directors of Guidance or Counseling (included with Attachment A) recently sent with these materials includes the instruction: "Another change in the ATP for 1981-82 is that Taking the SAT contains a new sample test. Please be sure that all students who register for the SAT receive a copy of this publication as far in advance of the test date as possible so that they have ample time to become familiar with the test."

Unlike Mr. Galligan, the majority of students who take the SAT have elected to take the Preliminary Scholastic Aptitude Test (PSAT) in October of their junior year. This test is a slightly abbreviated version of the SAT and acquaints the student with the test questions and format and provides a true-to-life practice testing situation. High school counselors provide to interested students the PSAT/NMSQT Student Bulletin, which includes a full-length test with answers and scoring directions. PSAT test-takers now automatically receive through their high schools a copy of the test questions, the correct answers, and their own answers at the time their scores are reported. They can then use this information in preparation for taking the SAT.

(3) A reprint of an article which appeared in last month's American School Board Journal, entitled "Standardized Testing Has Become Education's Latest Scapegoat." (Attachment B) This is a short article which provides some basic information about tests, including types of tests, how to take them, how to prepare for them and who develops them. I think the "crash course on test design" included at the end may be particularly helpful as an introduction to tests for new members of the committee and their staffs.

(4) A short brochure entitled, "Developing a Test." (Attachment C) This document describes very briefly the lengthy and elaborate 19-month cycle followed in the development of an ETS test. This brochure should also be helpful to newcomers to the field and will point out the meticulous care taken to identify defective questions throughout the developmental process.

(5) A document entitled, "Questioning the Questions" (Attachment D), which describes the process for challenging a question on an ETS test, from initial inquiry through appeal. Attached to this document are two charts which indicate the frequency of the different types of inquiries or challenges to ETS concerning questions on the six admissions and guidance testing programs administered by ETS, that are covered by the New York State law. The first chart shows the numbers of inquiries about test questions on all forms of the test (not just disclosed tests) from January 1, 1980 through June 1981. This table points out that the large majority of inquiries is made by test-takers immediately or shortly after the test administration, but prior to receipt of test questions and answers on disclosed tests. Only a very small portion of the 1980-81 inquiries were received as a result of test disclosure. In all, ten questions that were faulty or contained misprints were discovered; seven of them early enough so that they were not counted in scoring. The second chart deals only with disclosed tests and shows, for each of the six testing programs, how many tests have been disclosed, the number of questions disclosed, the number of questions challenged (after test-takers' receipt of questions and answers), the number of flawed questions, and the disposition made. The percentage of flawed items identified after disclosure is only five one-hundredths of one percent.

(6) A chart which indicates the numbers and names of states whose legislatures have introduced bills to regulate and disclose admissions or occupational tests. (Attachment E) This chart shows that there has been a decreased level of legislative activity in this area in 1981 compared to previous years, and that only one state has enacted a test disclosure law.

Finally, I also respectfully request that this letter be included as a part of the hearing record.

I thank you very much for your attention.

Sincerely,

Marion G. Epstein

Marion G. Epstein
Senior Adviser
Educational Testing Service

MGE:tgm
Attachments

cc: The Honorable Carl D. Perkins
The Honorable Thomas E. Coleman
The Honorable William F. Goodling

Domestic Shipments (Non-NY)

ADMISSIONS TESTING PROGRAM OF THE COLLEGE BOARD
BOX 592 • PRINCETON NJ 08541 • (609) ~~683-8500~~ 771-7600

To: Directors of Guidance or Counseling

Subject: Information about the Admissions Testing
Program (ATP) for 1981-82

This shipment of ATP materials includes copies of a Special Announcement to students about the College Board's SAT Question-and-Answer Service which will be introduced in 1981-82. This new service offers students the opportunity to order copies of their SAT test questions and answers if they take the SAT on the dates specified in the Special Announcement.

Notices on the front covers of the ATP Student Bulletin and Taking the SAT refer students to the Special Announcement for more information about the SAT Question-and-Answer Service. Please display this Special Announcement in a prominent place in your guidance office along with the ATP Student Bulletins so that each student who receives a copy of the Bulletin and Taking the SAT also obtains the Announcement.

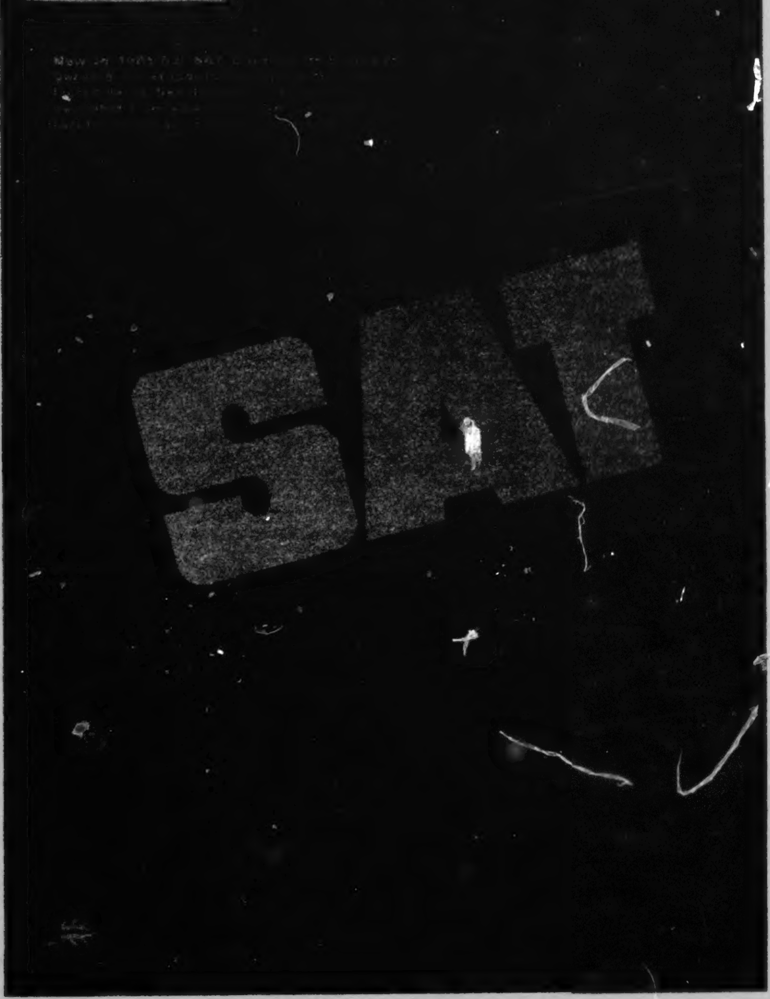
Another change in the ATP for 1981-82 is that Taking the SAT contains a new sample test. Please be sure that all students who register for the SAT receive a copy of this publication as far in advance of the test date as possible so that they have ample time to become familiar with the test.

The College Board will publish this September a compendium containing the five SATs made public during 1980-81. This booklet, 5 SATs, may be ordered for $5.00 each ($2.00 each for orders of 50 or more) by using the special order forms included in this shipment. Please note that 5 SATs is not intended to replace Taking the SAT, but rather is being made available to students and others who may be interested in seeing additional sample test questions.

The ATP Student Bulletin and Taking the SAT are being sent to you in separate shipments during the next few weeks. The quantity is based on the total number of copies you received last year and is noted on the Publications Shipment Notice enclosed with this shipment. If you need to reorder these or the Special Announcement about the SAT Question-and Answer Service during the year, please use the second copy of this form (Publications Reorder Form). Additional order forms for 5 SATs can be obtained upon request.

Thank you for your cooperation.

Attachment A



The Scholastic Aptitude Test 3
The Test of Standard Written English 3
Preparing for the Test 3
SAT-Verbal Sections 4
SAT-Mathematical Sections 8
The Test of Standard Written English 16
The Sample Test 19
How to Score Your Sample Test 44
Receiving Your Score Report 46
Answer Sheet 47
Six Points about Special Preparation for the SAT back cover

SAT Committee 1980-81

Jeremy [illegible]ilpatrick, University of Georgia, Athens, Georgia, *Chairman*
He[illegible]y F. Bedford, Phillips Exeter Academy, Exeter, New Hampshire
[illegible]ncy S. Cole, University of Pittsburgh, Pittsburgh, Pennsylvania
B[illegible]bara W. Davis, Charlotte-Mecklenburg Schools, Charlotte, North Carolina
Jeanette B. Hersey, Connecticut College, New London, Connecticut
Willie May, Wendell Phillips High School, Chicago, Illinois
Allen Parducci, University of California—Los Angeles, Los Angeles, California
Alonzo F. Rodriguez, Metropolitan State College, Denver, Colorado
Emory A. Sigler Jr., Highland Park High School, Dallas, Texas
Nathaniel Teich, University of Oregon, Eugene, Oregon

TSWE Committee 1980-81

Ronald B. Newman, University of Miami, Coral Gables, Florida, *Chairman*
Diana Culbertson, Kent State University, Kent, Ohio
Carlota Cárdenas de Dwyer, University of Texas, Austin, Texas
Maurice Lee, Vassar College, Poughkeepsie, New York
Marjorie G. Roemer, Brookline High School, Brookline, Massachusetts

---

The Admissions Testing Program is a program of the College Board, a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

This booklet was prepared and produced by Educational Testing Service (ETS), which develops and administers the tests of the Admissions Testing Program for the College Board.

Copies of the booklet are supplied to secondary schools for distribution free of charge to students who plan to register for the SAT. Individual copies may be ordered at $3 each (or in quantities of 50 or more at $2 each) from College Board Publication Orders, Box 2815, Princeton, NJ 08541.

Copyright © 1981 by College Entrance Examination Board. All rights reserved.

The sample test questions in this booklet are reprinted by permission of ETS, the copyright owner.

The College Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

The purpose of this booklet is to provide information that will help you take the Scholastic Aptitude Test (SAT) and the Test of Standard Written English (TSWE) with greater confidence.

Taking the SAT includes descriptions of the SAT and TSWE, an explanation of the different types of questions, and suggestions about how to approach answering the questions. In addition, the booklet contains hints for using your time efficiently and for guessing. Finally, a complete sample test and answer sheet are included as well as an explanation of how the answer sheets are scored.

If you have not already received a copy of this year's Admissions Testing Program *Student Bulletin*, ask for one at your school. The *Bulletin* tells how to register for the test, where and when to take it, and how your scores will be reported.

THE SCHOLASTIC APTITUDE TEST

The SAT is a multiple-choice test made up of separately timed verbal and mathematical sections. Verbal questions measure your ability to understand what you read and the extent of your vocabulary. Mathematical questions measure your ability to solve problems involving arithmetic reasoning, algebra, and geometry. These abilities have been shown to be related to successful academic performance in college. The SAT does not measure other kinds of abilities which may be associated with success in college, such as special talents or motivation.

Your high school record is probably the best evidence of your preparation for college. Because applicants have taken different courses and come from high schools with different grading practices, college admissions officers need a common measure of ability, such as the SAT. However, scores on the SAT are just part of the information used in making an admissions decision.

THE TEST OF STANDARD WRITTEN ENGLISH

The TSWE is a multiple-choice test administered with the SAT. The questions measure your ability to recognize standard written English, the language of most college textbooks and the English you will probably be expected to use in the papers you write in college. The TSWE is not intended to be used as an admissions test. Instead, after you have been admitted to college, TSWE scores may be used by the college to choose an English course that is best for you.

PREPARING FOR THE TEST

The abilities measured by the SAT develop over years of study and use, both in and out of school. For a statement about special preparation for the SAT, see the back cover of this booklet.

Studying the sample questions, directions, and explanations in this booklet will help you prepare for the test. The directions for the questions and the examples are printed here as they appear in the test book. If you study them now, you will understand them when you take the test.

The total testing time is three hours. Each test book is divided into six 30-minute sections:

2 SAT-verbal sections (a total of 85 questions)
2 SAT-mathematical sections (a total of 60 questions)
1 Test of Standard Written English (50 questions)
1 section of experimental verbal, mathematical, or TSWE questions, which do not count toward your score

The questions in the experimental section are used to maintain the quality of the tests and to provide essential information for future editions. The arrangement of sections is not the same in every edition of the test.

Within each group of questions of the same type, the questions are arranged more or less according to difficulty. The easier questions are usually at the beginning of the group and the more difficult at the end. Therefore, if you are working through a set of questions of a particular type and you find that the questions are becoming difficult for you, simply read quickly through the remaining questions in the group, answering only those that you feel you know. You should then begin working on the next group of questions. You can return to the questions you omitted if you have time. Remember, you get just as much credit for correctly answering easy questions as for correctly answering hard ones.

You get one point for each question you answer correctly, and you lose a fraction of a point for each question you answer incorrectly. You neither gain nor lose points for questions you omit. (See page 44 for scoring information.)

You may do scratchwork in the test book or mark questions you omitted and may want to go back to. You should *not* make extra marks on the answer sheet, because they could be misread as answers by the scoring machine.

Because of the way the test is scored, haphazard or random guessing for questions you know nothing about is unlikely to change your score. When you know that one or more choices can be eliminated, guessing from among the remaining choices should be to your advantage.

Many students who do well on the SAT omit some questions. Many students who receive average or slightly above average scores answer from 40 to 60 percent of the questions correctly. Questions vary in difficulty. Some are answered correctly by as many as 90 percent of the students, but some are answered correctly by as few as 10 percent.

Suggestions for answering specific types of questions follow.

SAT-VERBAL SECTIONS

There are four types of questions in the verbal sections of the SAT: 25 antonyms, 20 analogies, 15 sentence completions, and 25 questions based on reading passages. The antonyms usually take the least time, followed by analogies, sentence completion questions, and, finally, the reading comprehension questions. High-scoring students can do between 2 and 3 antonyms a minute, but the same students may take more than 7 minutes to read a 400-word passage and answer 5 questions on it.

Your answers to the 85 questions in the verbal sections make up your total verbal score. (See page 44.) The score report you'll receive will also show two subscores: (1) a vocabulary subscore, based on the antonym and analogy questions; (2) a reading subscore, based on the sentence completions and the questions on reading passages.

A careful balance of reading materials and words drawn from a variety of subject-matter fields helps assure that the test is fair to students with different interests. However, no specialized knowledge in science, social sciences, literature, or other fields is required.

Antonyms (opposites)

Antonym questions test the extent of your vocabulary. The vocabulary used in the antonym questions includes words that high school students come across in their general reading, although some words may not be the kind that you use in everyday speech.

Directions: Each question below consists of a word in capital letters, followed by five lettered words or phrases. Choose the word or phrase that is most nearly opposite in meaning to the word in capital letters. Since some of the questions require you to distinguish fine shades of meaning, consider all the choices before deciding which is best.

EXAMPLE:

GOOD: (A) sour (B) bad (C) red
(D) hot (E) ugly

Ⓐ●ⒸⒹⒺ

You can probably answer this example without carefully considering all of the choices. However, most of the antonyms in the verbal section require more careful analysis. When you work on antonym questions, remember that:

1. Among the five choices offered, you are looking for the word that means the *opposite* of the given word. No words that have the same meaning as the given word are included among the five choices.

2. You are looking for the *best* answer. Read all of the choices before deciding which one is best, even if you feel sure you know the answer. For example:

   SUBSEQUENT: (A) primary (B) recent
   (C) contemporary (D) prior (E) simultaneous

   Someone working quickly might choose (B) *recent* because it refers to a past action and *subsequent* refers to a future action. However, choice (D) *prior* is the best answer.

3. Few words have exact opposites, that is, words that are opposite in all of their meanings. You should find the word that is *most nearly* opposite. For example:

   FERMENTING: (A) improvising (B) stagnating
   (C) wavering (D) plunging deeply (E) dissolving

   Even though *fermenting* is normally associated with chemical reactions, whereas *stagnating* is normally associated with water, *fermenting* means being agitated and *stagnating* means being motionless. Therefore, choice (B) *stagnating* is the best antonym of *fermenting*.

4. You need to be flexible. A word can have several meanings. For example:

   DEPRESS: (A) force (B) allow (C) clarify
   (D) elate (E) loosen

   The word *depress* can mean "to push down." However, no word meaning "to lift up" is included among the choices. Therefore, you must consider another meaning of *depress*, "to sadden or discourage." Option (D) *elate* means to fill with joy or pride. The best answer is (D) *elate*.

5. You could try to put the word in a sentence. If you don't know the dictionary meaning of a word but have a feeling for how the word should be used, try to make up a short phrase or sentence using the word. This may give you a clue as to which choice is

an opposite, even though you may not be able to define the word precisely.

INCUMBENT: (A) conscious (B) effortless
(C) optional (D) improper (E) irrelevant

One of the meanings of *incumbent*, and the one you may know best, is "the holder of an office" as in the sentence, "The incumbent president usually has a better chance of winning an election than the challenger." However, an opposite of that meaning is not included in the choices. Try to think of another use of *incumbent*, such as in the sentence, "It is incumbent upon me to finish this." If you can think of such a phrase, you realize that *incumbent* means "imposed as a duty" or "obligatory."

Of the five choices, (A), (B), and (D) are in no way opposites of *incumbent* and you can easily eliminate them. Choice (E) means "not pertinent" and choice (C) means "not compulsory." Although choice (E) may look attractive, choice (C) *optional* is more nearly an exact opposite to *incumbent*. Choice (C), therefore, is the answer.

6. Because answering antonyms depends on knowing the meanings and uses of words, memorizing word lists is probably of little use. Anything that helps you to think about words and how they are used improves your verbal ability and is likely to improve your performance on antonyms and other kinds of verbal questions. Read carefully some moderately complex books or some good magazines on topics with which you are not familiar. If you come across unfamiliar words that you can't understand from the context, use a dictionary. Do crossword puzzles or play word games with your friends or family. These exercises should be at least as helpful as studying word lists and certainly more interesting.

Analogies

Analogy questions test your ability to see a relationship in a pair of words, to understand the ideas expressed in the relationship, and to recognize a similar or parallel relationship.

Directions: **Each question below consists of a related pair of words or phrases, followed by five lettered pairs of words or phrases. Select the lettered pair that best expresses a relationship similar to that expressed in the original pair.**

EXAMPLE:
YAWN : BOREDOM : : (A) dream : sleep
(B) anger : madness (C) smile : amusement
(D) face : expression
(E) impatience : rebellion
Ⓐ Ⓑ ● Ⓓ Ⓔ

When you answer analogy questions,

1. Try to establish a precise relationship between the first two words before examining the answer choices. State this relationship as clearly as you can in a sentence or phrase. Next, find the choice with a pair of words that have the same relationship, and that express the same idea as the original pair of words. In the example above, a yawn is a sign of boredom in the same way that a smile is a sign of amusement. Another example is the following:

SUBMISSIVE : LED : : (A) wealthy : employed
(B) intolerant : indulged (C) humble : humiliated
(D) incorrigible : taught
(E) inconspicuous : overlooked

The relationship between *submissive* and *led* could be expressed as "to be submissive is to be easily led." Only choice (E) has the same relationship as *submissive* and *led*: "to be inconspicuous is to be easily overlooked" parallels "to be submissive is to be easily led."

Although the wealthy may find it easier to get employment than do the poor, the statement "to be wealthy is to be easily employed" is an expression of opinion and not an expression of the relationship between the words according to their dictionary meanings. Remember that the relationship that is to be established between the two words should take into account either the dictionary or implied meanings of the words.

2. Practice recognizing relationships. Below are three examples of the relationships that could be used.

SONG : REPERTOIRE : : (A) score : melody
(B) instrument : artist (C) solo : chorus
(D) benediction : church (E) suit : wardrobe

The best answer is choice (E). The relationship between the words can be expressed as "several (first word) make up a (second word)." Several (songs) make up a (repertoire) as several (suits) make up a (wardrobe).

REQUEST : ENTREAT : : (A) control : explode
(B) admire : idolize (C) borrow : steal
(D) repeat : plead (E) cancel : invalidate

The best answer is choice (B). Although both words have similar meanings, they express different degrees of feeling. To (entreat) is to (request) with strong feeling as to (idolize) is to (admire) with strong feeling.

To answer analogy questions, you must think carefully about the precise meanings of words. For instance, if you thought the word "entreat" meant only "to ask" instead of "to ask urgently," you would have trouble establishing the correct relationship between *request* and *entreat*.

FAMINE : STARVATION : : (A) deluge : flood
(B) drought : vegetation (C) war : treaty
(D) success : achievement (E) seed : mutation

The best answer is choice (A). (Famine) results in (starvation) as a (deluge) results in a (flood).

3. Don't be misled by relationships that are close to but not parallel to the relationship in the original pair. All of the pairs of words listed in the choices have relationships that can be stated; however, the correct answer has most nearly the same relationship as the original pair. Look at the following example.

   KNIFE : INCISION : : (A) bulldozer : excavation
   (B) tool : operation (C) pencil : calculation
   (D) hose : irrigation (E) plow : agriculture

   On the most general level, the relationship between *knife* and *incision* is that the object indicated by the first word is used to perform the action indicated by the second word. Since "a knife is used to make an (incision)," "a bulldozer is used to make an (excavation)," and "a hose is used for (irrigation)," there appear to be two correct answers. You need to go back and state the relationship more precisely. Some aspect of the relationship between the original pair exists in only one of the choices. A more precise relationship between *knife* and *incision* could be expressed as: "a knife cuts into something to make an incision" and "a bulldozer cuts into something to make an excavation." This relationship eliminates *hose : irrigation* as a possible answer, and no other relationship between *hose : irrigation* parallels the relationship to *knife : incision* as well. The best answer is choice (A).

4. Remember that words can have more than one relationship. For example:

   PRIDE : LION : : (A) snake : python (B) pack : wolf
   (C) rat : mouse (D) bird : starling (E) dog : canine

   A possible relationship between *pride* and *lion* might be that "the first term describes a characteristic of the second (especially in mythology)." Using this reasoning, you might look for an answer such as *wisdom : owl*, but none of the given choices has that kind of relationship. Another relationship between *pride* and *lion* is "a group of lions is called a pride"; therefore, the answer is (B) *pack : wolf;* "a group of wolves is called a pack."

Sentence Completion Questions

Sentence completion questions test your ability to recognize the relationships among parts of a sentence. You are given a sentence from which one or two words have been removed and asked to complete the sentence by choosing the word or words that are consistent with other parts. Sentence completion questions ask you to know the words listed as choices and their proper use in the context of a sentence. The sentences, taken from published material, cover a variety of topics. Each sentence provides enough information so that you can find the correct answer without any information beyond what is contained in the sentence itself.

Directions: Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five lettered words or sets of words. Choose the word or set of words that *best* fits the meaning of the sentence as a whole.

EXAMPLE:
Although its publicity has been ---, the film itself is intelligent, well-acted, handsomely produced, and altogether ---.

(A) tasteless . . respectable (B) extensive . . moderate
(C) sophisticated . . amateur (D) risqué . . crude
(E) perfect . . spectacular

● Ⓑ Ⓒ Ⓓ Ⓔ

The word *although* suggests that the publicity gave the wrong impression of the movie, so look for two words that are more or less opposite in meaning. Also, the second word has to fit in with "intelligent, well-acted, handsomely produced." Choices (B), (D), and (E) are not opposites. Choice (C) can't be the correct answer even though *sophisticated* and *amateur* are nearly opposites, because an "intelligent, well-acted, handsomely produced" film isn't amateurish. Only choice (A), when inserted in the sentence, gives a logical statement.

Here are some suggestions to keep in mind when you answer sentence completion questions:

1. Read the entire sentence carefully; make sure you understand the ideas being expressed.

2. Don't select an answer simply because it is a popular cliché or "sounds good."

3. Look for grammatical clues within the sentence. For example:

   The excitement does not --- but --- his senses, giving him a keener perception of a thousand details.
   (A) slow . . diverts (B) blur . . sharpens
   (C) overrule . . constricts (D) heighten . . aggravates
   (E) forewarn . . quickens

   The word *but* implies that the answer will involve two words that are more or less opposite in meaning. If you keep this in mind, you can eliminate all of the choices except for (B) *blur . . sharpens*. Only the words in choice (B) imply opposition. Also, "sharpens his senses" is consistent with the notion that he had a "keener perception of a thousand details."

4. If the sentence has two blanks to be filled, make sure that *both* words make sense in the sentence. For example:

   They argue that the author was determined to --- his own conclusion, so he --- any information that did not support it.

   (A) uphold . . ignored (B) revise. . . destroyed
   (C) advance . . devised (D) disprove . . distorted
   (E) reverse . . confiscated

   The first words in choices (A) *uphold . . ignored* and (C) *advance . . devised* seem all right. However, the second word in choice (C) *advance . . devised* does not make sense in the sentence; why would an author who wished to advance his theory devise information that did not support it? Only choice (A) makes a logically consistent sentence.

Mr. Dillon is a skeptic, — to believe that the accepted opinion of the majority is generally —.

(A) prone . . infallible (B) afraid . . misleading
(C) inclined . . justifiable (D) quick . . significant
(E) disposed . . erroneous

The words to be inserted in the blank spaces in the question above must result in a statement that is consistent with the definition of a skeptic. Since a skeptic would hardly consider the accepted opinion of the majority as *infallible, justifiable*, or *significant*, you can eliminate choices (A), (C), and (D). A skeptic would not be *afraid* that the accepted opinion of the majority is *misleading;* he would believe that it was. Therefore, choice (B) is not correct. Only choice (E) *disposed . . erroneous* yields a logical sentence.

5. After choosing an answer, read the entire sentence to yourself. Make sure that the sentence is logical.

Reading Passages

The reading comprehension section of the SAT presents a fairly familiar task, especially if you have read widely. The reading comprehension section tests a variety of skills. Some questions test your understanding of what is stated directly. Sometimes you have to interpret and analyze what you read. Some questions test your ability to recognize applications of the author's principles or opinions. Others require you to judge what you have read—to identify strong and weak points in the presentation, to determine how well the author supports any claims with evidence, and to recognize the means by which the author communicates points. Each test contains passages drawn from the following categories:

Narrative:	(novels, short stories, biographies, essays)
Biological Science:	(medicine, botany, zoology)
Physical Science:	(chemistry, physics, astronomy)
Humanities:	(art, literature, music, philosophy, folklore)
Social Studies:	(history, economics, sociology, government)
Argumentative:	(the presentation of a definite point of view on some subject)

Each passage contains all the information you need to answer a question. However, passages about subjects you are familiar with or interested in may be easier for you. If you find a passage that seems difficult to you, skip it and go on. You will be omitting only a few questions and you will be saving yourself time that can be better used somewhere else.

Is it better to read the questions before reading the passage? Probably not, but the content of a passage may help you decide. For instance, if you're good in science, you may follow the reasoning in a science passage and remember most of the details. Looking at the questions before you read the science passage would probably be a waste of time for you. However, if you find science passages confusing, reading the questions first may be worth your time. While you are doing the sample test, try both methods and see which works better for you in terms of selecting the correct answer quickly.

Read closely and attentively. Follow the reasoning; notice how each piece of information relates to the ideas being presented. Notice attitudes, tone, and general style. You may want to check an important fact or idea, but don't waste time underlining or making notes in the margin. Try to get a sense of the principal ideas, facts, and organization of the passage.

Here are some suggestions that may be helpful:

1. Read all of the choices before you select an answer.
2. Answer questions on the basis of what is in the passage. Do not answer questions on the basis of your personal opinion or knowledge.
3. Answer the question that is asked. Do not pick one of the choices simply because it is a true statement.
4. Make sure that the answer you choose is completely correct. Do not be misled by choices that are partially true.

Several types of questions are asked about the passages. Some questions test your ability to differentiate between the main idea of a passage and secondary and supporting information. In answering questions that ask for the main idea, don't be distracted by statements that are true according to the passage but that are secondary or supportive to the central point. Below is a short excerpt from a passage that presents a particular point of view of Plato's *Republic*. The author of this passage expresses ideas strongly so the passage would be classified as argumentative.

Directions: The passage below is followed by questions based on its content. Answer all questions following the passage on the basis of what is *stated* or *implied* in that passage.

That Plato's *Republic* should have been admired, on its political side by decent people, is perhaps the most astonishing example of literary snobbery in all history.

Let us consider a few points in this totalitarian tract. The main purpose of education is to produce courage in battle. To this end, there is to be a rigid censorship of the stories told by mothers and nurses to young children; there is to be no reading of Homer because that degraded versifier makes

heroes lament and gods laugh; the drama is to be forbidden because it contains villains and women; music is to be only of certain kinds, which, in modern terms, would be military bands playing "My Country Tis of Thee" and "Stars and Stripes Forever."

A *main idea* question that might be asked is:

1. The main point of the passage is to

- **(A) cast contempt on the kind of music advocated in Plato's *Republic***
- **(B) describe the content of Plato's *Republic***
- **(C) discuss the positive and negative aspects of Plato's *Republic***
- **(D) show how Plato's *Republic* influenced the lives of people at the time**
- **(E) criticize the political philosophy contained in Plato's *Republic***

Another *main idea* question that might be asked is:

2. Which of the following would be the most appropriate title for the passage?

- **(A) The Perfection of Plato's *Republic***
- **(B) Why Plato's *Republic* Has Been Censored**
- **(C) Plato's *Republic*: A Totalitarian Tract**
- **(D) The Heroes of Plato's *Republic***
- **(E) Plato: *The Moralist***

Both questions require only that you understand the main point, which the author makes in the first two sentences; the rest of the passage gives supporting detail. Note that the main point of a passage is not always stated at the beginning. In question 1, the statement in choice (A) mentions one of the supporting points, not the main idea. Choice (B) gives some of the content of the passage, but the main thrust of the passage is much more negative than the statement in choice (B). Choice (C) is inaccurate because the author doesn't admit that the book has any positive aspects. Choice (D) is not supported. Only choice (E) adequately describes the main idea. For the second question, follow the kind of analysis made of question 1. You'll see that only option (C) accurately reflects the main idea of the passage.

Another type of question asks for details stated in the passage. Once again, be sure to answer the question that is asked. For example:

3. According to the passage, which of the following are statements of Plato's beliefs?

- **I. Drama should expose the weaknesses of villains.**
- **II. Only those parts of Homer dealing with the heroes and gods may be used.**
- **III. Stories told to children should be strongly censored.**

(A) I only (B) II only (C) III only
(D) II and III only (E) I, II, and III

Statement I is inaccurate because no drama is permitted, and statement II is inaccurate as no parts of Homer may be read. Only statement III is accurate and the correct answer is choice (C).

You can use partial information to help you, even if you're not sure of the answer. In the question above, if you know that statement I is incorrect, you can eliminate choices (A) and (E), leaving only three other answers to choose from. If you are doubtful about II but sure of III, you can eliminate (A), (E), and (B), leaving you one chance in two of choosing correctly between (C) and (D).

4. The passage suggests that all of the following are forbidden in Plato's *Republic* EXCEPT

- **(A) dance music**
- **(B) patriotic music**
- **(C) plays about villains**
- **(D) portrayals of the amusements of the gods**
- **(E) stories about poor people stealing to feed their families**

Question 4 asks you to identify which of the choices is not forbidden. Since the author mentions only that songs played by military bands are permitted, you must recognize that the suggestion is that patriotic music is allowed. This is reinforced by the connotations of the titles of the songs given as examples. The correct answer is option (B).

Some of the questions ask you to evaluate or interpret the passage. Following is an example of such a question:

5. The author's attitude toward Plato's *Republic* is one of

(A) quiet concern (B) cautious acceptance
(C) reverent admiration (D) outraged disapproval
(E) total indifference

Question 5 requires you to identify the author's attitude toward Plato's philosophy. A strong attitude of disapproval is evident in almost every sentence, but most obviously in line 4 in which the author refers to Plato's *Republic* as a "totalitarian tract." The correct answer is (D).

You can improve your reading comprehension by reading extensively and, what is most important, thinking while you are reading. If you have limited your reading to only one or two particular topics, practice reading about unfamiliar subjects. Many books and good science, news, literature, and art magazines provide interesting material. While you are reading, follow the author's argument, line of reasoning, methods used to make a point, and the implications of what is said.

SAT-MATHEMATICAL SECTIONS

Some questions in the mathematical sections of the SAT require you to apply numerical, graphic, spatial, symbolic, and logical techniques to situations familiar to you; these may be similar to exercises in your textbooks. Other questions may require you to do some original thinking. The mathematical preparation expected is a year of algebra and some geometry. Although most of the geometric ideas involved are usually taught in the elementary and junior high years, a few of the questions involve topics that are first taught in high school geometry. Many of these geometric concepts are reviewed in the material that follows.

Two types of multiple-choice questions are used in the SAT-mathematical sections:

1. Standard multiple-choice questions (approximately two-thirds of the questions)
2. Quantitative comparison questions (approximately one-third of the questions)

These types will be explained later. Most of the questions are classified as arithmetic, algebra, or geometry, and there is approximately the same number of each type.

Some Mathematical Concepts with Which You Should Be Familiar

Arithmetic—basic addition, subtraction, multiplication, and division; percent; average; odd and even numbers; prime numbers; divisibility (for example, 24 is divisible by 8)

Algebra—negative numbers; simplifying algebraic expressions; linear equations; inequalities; simple quadratic equations; positive integer exponents; roots

Geometry—area (square, rectangle, triangle, and circle); perimeter of a polygon; circumference of a circle; volume of a box and cube; special properties of isosceles, equilateral, and right triangles; 30°—60°—90° and 45°—45°—90° triangles; properties of parallel and perpendicular lines; locating points on a coordinate grid

This section will help you review or refresh your knowledge about mathematical vocabulary and concepts frequently needed to solve problems.

Words and Phrases You Should Know

When You See:	Think:
Positive Integers	**1, 2, 3, 4, . . .**
Negative Integers	**$-1, -2, -3, -4, \ldots$**
Integers	**$\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots$**
Odd Numbers	**1, 3, 5, 7, 9, . . .**
Even Numbers	**0, 2, 4, 6, 8, . . .**
Consecutive Integers	**$n, n+1, n+2, \ldots$ (n = an integer)**
Prime Numbers	**2, 3, 5, 7, 11, 13, 17, 19, . . .**
Average	**The sum of terms divided by the number of terms, for example: Average of 9, 11, and 16 $= \frac{9+11+16}{3} = 12$**

Concepts You Should Know

Odd and Even Numbers

Addition:

even + even = even
odd + odd = even
even + odd = odd

Multiplication:

even × even = even
even × odd = even
odd × odd = odd

Percent

Percent means hundredths or number out of 100, so that $\frac{40}{100}$ = 40 percent and 3 is 75 percent of 4 (because $\frac{3}{4} = \frac{75}{100}$ = 75 percent).

Some Percent Equivalents

$\frac{1}{10} = 0.1 = 10\%$

$\frac{1}{5} = 0.2 = 20\%$

$\frac{1}{2} = 0.5 = 50\%$

$\frac{1}{1} = 1.0 = 100\%$

$\frac{2}{1} = 2.0 = 200\%$

Note: To convert a fraction or decimal to percent, multiply by 100.

General Method of Converting a Fraction $\frac{a}{b}$ to a Percent:

$\frac{a}{b} = \frac{x}{100}$

$x = 100\left(\frac{a}{b}\right)$

Example: $\frac{3}{4} = \frac{x}{100}$

Therefore, $x = 100\left(\frac{3}{4}\right) = 75$

$\frac{3}{4} = \frac{75}{100} = 75\%$

Percents Greater Than 100

Problem: 5 is what percent of 2?

Solution: $\frac{5}{2} = \frac{x}{100}$

$x = \frac{500}{2} = 250$

Therefore, $\frac{5}{2} = \frac{250}{100} = 250\%$

So, 5 is 250 percent of 2. Note that this is equivalent to saying that 5 is $2\frac{1}{2}$ times 2.

Problem: Sue earned \$10 on Monday and \$12 on Tuesday. The amount earned on Tuesday was what percent of the amount earned on Monday?

An equivalent question is "\$12 is what percent of \$10?"

Solution: $\frac{12}{10} = \frac{x}{100}$

$x = \frac{1{,}200}{10} = 120$

So, $\frac{12}{10} = \frac{120}{100} = 120\%$

Percents Less Than 1

Problem: 3 is what percent of 1,000?

Solution: $\frac{3}{1,000} = 0.003 = 0.3\%$ or $\frac{3}{10}$ of 1 percent

Problem: Socks are $1.00 a pair or 2 pairs for $1.99. The savings in buying 2 pairs is what percent of the total cost at the single pair rate?

Solution: At the single pair rate, 2 pairs would cost $2.00, so the savings is only $0.01. Therefore, you must answer the question "$0.01 is what percent of $2.00?"

Because $\frac{0.01}{2.00} = \frac{0.5}{100}$, the savings is 0.5% or $\frac{1}{2}$ of 1 percent.

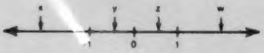
Squares of Integers

n	1	2	3	4	5	6	7	8	9	10	11	12
n^2	1	4	9	16	25	36	49	64	81	100	121	144

n	−1	−2	−3	−4	−5	−6	−7	−8	−9	−10	−11	−12
n^2	1	4	9	16	25	36	49	64	81	100	121	144

Signed Number Properties

positive × positive = positive
negative × negative = positive
negative × positive = negative
$-(a - b) = b - a$
$(-x)^2 = x^2$
If $x < 0$, $x^2 > 0$



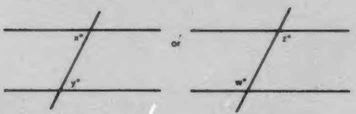
On the number line above:
$x < y$ For example, $-2 < -\frac{1}{2}$
$y^2 > 0$
$z^2 < z$ For example, $\left(\frac{1}{2}\right)^2 < \frac{1}{2}$
$x^2 > z$ For example, $(-2)^2 > \frac{1}{2}$
$z^2 < w$
$x + z < 0$
$y - x > 0$

Factoring

$x^2 + 2x = x(x + 2)$
$x^2 - 1 = (x + 1)(x - 1)$
$x^2 + 2x + 1 = (x + 1)(x + 1) = (x + 1)^2$
$x^2 - 3x - 4 = (x - 4)(x + 1)$

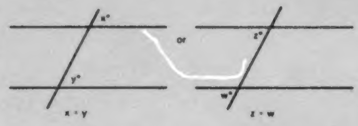
Properties of Parallel Lines

1. If two parallel lines are cut by a third line, the alternate interior angles are equal.
   For example:



$x = y$ $z = w$

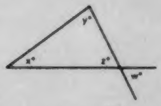
2. If two parallel lines are cut by a third line, the corresponding angles are equal.
   For example:



$x = y$ $z = w$

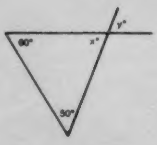
Note: Words like "alternate interior" or "corresponding" are generally not used on the test, but you do need to know which angles are equal.

Angle Relationships

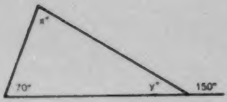


$x + y + z = 180$
(Because the sum of the interior angles of a triangle is 180°)

$z = w$
(When two straight lines intersect, vertical angles are equal.)

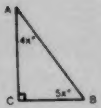


$y = 70$
(Because x is equal to y and $60 + 50 + x = 180$)

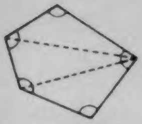


y = 30
(Because a straight angle is 180°, y = 180 − 150)

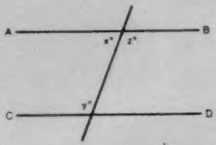
x = 80
(Because 70 + 30 + x = 180)



x = 10
(Because 4x + 5x = 90°) Also, the length of side AC is greater than the length of side BC (Because ∠ B is greater than ∠ A)

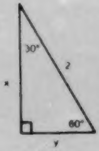


The sum of all angles of the polygon above is 3 (180°) = 540° because it can be divided into 3 triangles, each containing 180°.



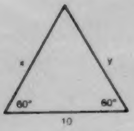
If AB is parallel to CD, then x + y = 180 (Because x + z = 180 and y = z)

Side Relationships

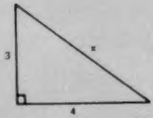


y = 1
(Because the length of the side opposite the 30° angle in a right triangle is half the length of the hypotenuse)

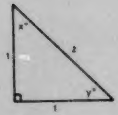
$x = \sqrt{3}$
(By the Pythagorean Theorem,
$x^2 + 1^2 = 2^2$
$x^2 = 3$
$x = \sqrt{3}$)



x = y = 10
(Because the unmarked angle is 60°, all angles of the triangle are equal, and, therefore, all sides of the triangle are equal)



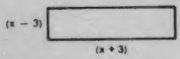
x = 5
(By the Pythagorean Theorem,
$x^2 = 3^2 + 4^2$
$x^2 = 9 + 16$
$x^2 = 25$
$x = \sqrt{25} = 5$)



x = y = 45°
(Because two sides are equal, the right triangle is isosceles and angles x and y are equal. Also, x + y = 90 which makes both angles 45°)
$z = \sqrt{2}$
(Because $1^2 + 1^2 = z^2$)

Area and Perimeter Formulas

Area of a rectangle = length × width = L × W
Perimeter of a rectangle = 2(L + W)
Examples:



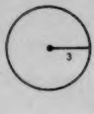
Area = 12

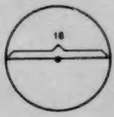
Perimeter = 14

Area = (x − 3) (x + 3) = $x^2 - 9$

Perimeter = 2 [(x + 3) + (x − 3)]
= 2 (2x) = 4x

Area of a circle = πr^2 (where r is the radius)
Circumference of a circle = $2\pi r = \pi d$ (where d is the diameter)
Examples:

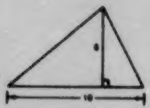


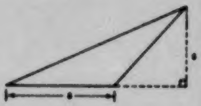


Area = $\pi(3^2) = 9\pi$
Circumference = $2\pi(3)$
$= 6\pi$

Area = $\pi(8^2) = 64\pi$
Circumference = $\pi(16) = 16\pi$

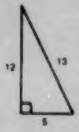
Area of a triangle = $\frac{1}{2}$(altitude × base) = $\frac{1}{2}a \cdot b$

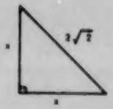




Area = $\frac{1}{2} \cdot 6 \cdot 10 = 30$

Area = $\frac{1}{2} \cdot 6 \cdot 8 = 24$





Area = $\frac{1}{2} \cdot 12 \cdot 5 = 30$

Perimeter = $12 + 5 + 13 = 30$

$x = 2$

(Because $x^2 + x^2 = (2\sqrt{2})^2$

$2x^2 = 4 \cdot 2$

$x^2 = 4$

$x = 2$)

Area = $\frac{1}{2} \cdot 2 \cdot 2 = 2$

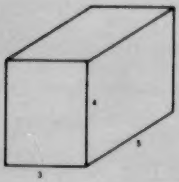
Perimeter = $2 + 2 + 2\sqrt{2}$

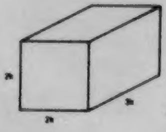
$= 4 + 2\sqrt{2}$

Volume of a Rectangular Solid (box)

Volume of a box = length × width × height = L · W · H

Examples:





Volume = $5 \cdot 3 \cdot 4 = 60$

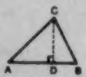
Volume = $(3h)(2h)(2h) = 12h^3$

The formulas and symbols given in the directions that follow appear in the test book. Learning them now will help you when you take the actual test.

Standard Multiple-Choice Questions

Directions: In this section solve each problem, using any available space on the page for scratchwork. Then decide which is the best of the choices given and blacken the corresponding space on the answer sheet. The following information is for your reference in solving some of the problems.

Circle of radius r:
Area = πr^2
Circumference = $2\pi r$
The number of degrees of arc in a circle is 360.
The measure in degrees of a straight angle is 180.



If ∠ CDA is a right angle, then (1) area of △ ABC = $\frac{AB \times CD}{2}$

(2) $AC^2 = AD^2 + DC^2$

Triangle: The sum of the measures in degrees of the angles of a triangle is 180.

Definition of symbols:

= is equal to	≦ is less than or equal to
≠ is unequal to	≧ is greater than or equal to
< is less than	∥ is parallel to
> is greater than	⊥ is perpendicular to

Note: Figures which accompany problems in this test are intended to provide information useful in solving the problems. They are drawn as accurately as possible EXCEPT when it is stated in a specific problem that its figure is not drawn to scale. All figures lie in a plane unless otherwise indicated. All numbers used are real numbers.

When you take the SAT, remember to use the available space in the test book for scratchwork. You are *not* expected to do all the reasoning and figuring in your head.

The problems that follow will give you an idea of the type of mathematical thinking required. First, try to answer each question yourself. Then read the explanation, which may give you new insights into solving the problem and perhaps point out techniques that you will be able to use again. Note that the directions indicate to select the *best* of the choices given.

1. If $2a + b = 5$, then $4a + 2b =$

(A) $\frac{5}{4}$ (B) $\frac{5}{2}$ (C) 10 (D) 20 (E) 25

This is an example of a problem which requires realizing that $4a + 2b = 2(2a + b)$. Therefore, $4a + 2b = 2(2a + b) = 2(5) = 10$. The correct answer is (C).

2. If $16 \cdot 16 \cdot 16 = 8 \cdot 8 \cdot P$, then $P =$

(A) 4 (B) 8 (C) 32 (D) 48 (E) 64

This question can be solved by several methods. A time-consuming method would be to multiply the three 16s and then divide the result by the product of 8 and 8. A quicker approach would be to find what additional factors are needed in the right side of the equation to match those in the left side. These are two 2s and a 16, the product of which is 64. Yet another method involves solving for P as follows:

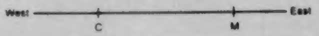
$$P = \frac{\overset{2}{\cancel{16}} \cdot \overset{2}{\cancel{16}} \cdot 16}{\cancel{8} \cdot \cancel{8}} = 2 \cdot 2 \cdot 16 = 64$$

The correct answer is (E).

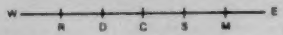
3. The town of Mason is located on Eagle Lake. The town of Canton is west of Mason. Sinclair is east of Canton, but west of Mason. Dexter is east of Richmond, but west of Sinclair and Canton. Assuming all these towns are in the United States, which town is farthest west?

(A) Mason (B) Dexter (C) Canton
(D) Sinclair (E) Richmond

For this kind of problem, drawing a diagram may help. In this case, a line can be effectively used to locate the relative position of each town. Start with the statement "The town of Canton is west of Mason" and, using abbreviations, draw the following:



From the remaining information, place the other towns in their correct order:



The final sketch shows that the town farthest west is Richmond (R) and the correct answer is (E).

4. If the average of seven x's is 7, what is the average of fourteen x's?

(A) $\frac{1}{7}$ (B) $\frac{1}{2}$ (C) 1 (D) 7 (E) 14

Don't get caught up in the wording of this problem which might lead you to choose (E) 14. The average of any number of equal numbers such as x is always x. Since you are given that the average of seven x's is 7, it follows that x = 7 and that the average of fourteen x's is also 7. The correct answer is (D).

5. If an asterisk (*) between two expressions indicates that the expression on the right exceeds the expression on the left by 1, which of the following is (are) true for all real numbers x?

I. $x(x + 2) * (x + 1)^2$
II. $x^2 * (x + 1)^2$
III. $\frac{x}{y} * \frac{x + 1}{y + 1}$

(A) None (B) I only (C) II only
(D) III only (E) I and III

This kind of problem involves working with a newly defined symbol. One approach is to check the statements one at a time. Statement I reduces to $x^2 + 2x * x^2 + 2x + 1$, so the expression on the right does exceed the expression on the left by 1. Therefore, statement I is true. Statement II reduces to $x^2 * x^2 + 2x + 1$ so the right expression exceeds the left expression by 2x + 1, which is not equal to 1 except when x = 0. This makes statement II *false*. Statement III is more difficult to check, but you can verify by subtraction or by substituting numbers (for example, x = 3, y = 5), that the expression on the right does not exceed the expression on the left by 1. Therefore, statement III is *false*. The only true statement is I, so the correct answer is (B).

In a problem of this kind, if you are able to decide about only one or two statements, you can still eliminate some choices and guess among those remaining. For example, if you can conclude that I is true and II is false, then the correct answer is either (B) or (E) because these choices contain statement I.

6. If a car travels X kilometers of a trip in H hours, in how many hours can it travel the next Y kilometers at this rate?

(A) $\frac{XY}{H}$ (B) $\frac{HY}{X}$ (C) $\frac{HX}{Y}$ (D) $\frac{H + Y}{X}$ (E) $\frac{X + Y}{H}$

You can solve this problem by using ratios or by using the distance formula.

Using the ratio method, X kilometers is to H hours as Y kilometers is to □ hours, where □ represents the amount of time required to travel Y kilometers:

$$\frac{X}{H} = \frac{Y}{\square}$$

$$X\square = HY$$

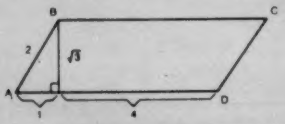
$$\square = \frac{HY}{X}$$

The correct answer is (B).

7. If 90 percent of P is 30 percent of Q, then Q is what percent of P?

(A) 3% (B) 27% (C) 30% (D) 270% (E) 300%

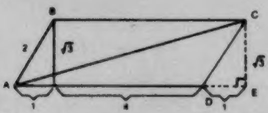
Writing an algebraic equation for this percent problem not only simplifies the work, but also helps you to organize your thoughts. "90 percent of P is 30 percent of Q" can be written as .90P = .30Q (or $\frac{9}{10}P = \frac{3}{10}Q$). "Q is what percent of P" tells you to find $\frac{Q}{P}$ and express it as a percent. $\frac{Q}{P} = 3$ and, therefore, Q is 300 percent of P and the correct answer is (E). (See pages 9-10 for review of percent.)



8. The figure above shows a piece of paper in the shape of a parallelogram with measurements as indicated. If the paper is tacked at its center to a flat surface and then rotated about its center, the points covered by the paper will be a circular region of diameter

(A) $\sqrt{3}$ (B) 2 (C) 5 (D) $\sqrt{28}$ (E) $\sqrt{30}$

The first step in solving the problem is to realize that the center of the parallelogram is the point of intersection of the two diagonals; thus, the diameter you are looking for is the length of the longer diagonal AC. One way to find AC is to think of the additional lines drawn as shown below.



The triangles at each end are congruent (equal in size and shape), so the length of DE and CE are 1 and $\sqrt{3}$, respectively. AEC is a right triangle; therefore, the Pythagorean Theorem can be used in solving the problem:

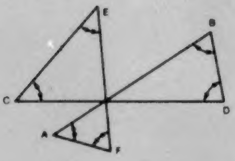
$AC^2 = CE^2 + AE^2$
$AC^2 = (\sqrt{3})^2 + (6)^2 = 3 + 36 = 39$

The diameter AC is $\sqrt{39}$ and the correct answer is (E).

9. An accurate 12-hour clock shows 3 o'clock at a certain instant. Exactly 11,999,999,995 hours later, what time does the clock show?

(A) 5 o'clock (B) 7 o[illegible]ck (C) 8 o'clock
(D) 9 o'clock (E) 10 o'clock

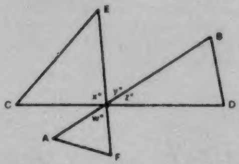
Every 12 hours the clock will again show 3 o'clock. You want to know the remainder when 11,999,999,995 is divided by 12 to see how many hours to add to 3 o'clock. Because this division problem will be cumbersome, think about a shortcut. Twelve divides 12,000,000,000 with no remainder and 11,999,999,995 is 5 less than 12,000,000,000. After 12,000,000,000, the clock will show 3 o'clock, so 5 hours before 12,000,000,000, the clock will show 10 o'clock. The correct answer is (E).



10. In the triangles above, if AB, CD, and EF are line segments, what is the sum of the measures of the six marked angles?

(A) 180° (B) 360° (C) 540° (D) 720°
(E) It cannot be determined from the information given.

This problem requires a creative problem-solving approach. One solution involves recognizing that the sum of the three unmarked angles in the triangles is 180. This can be seen from the following figure:



Because CD is a line segment, the sum of angles x, y, and z is 180°. Also, y = w because they are vertical angles. Therefore, x + w + z = 180. Since the sum of the measures of all angles in the three triangles is 540° (3 · 180°) and the sum of the unmarked angles of the triangles in the original figure equals 180°, it follows that the sum of the marked angles is 540° − 180° = 360°. The correct answer is (B). With this type of problem, if you do not reach a solution in a minute or so, go on to the next problem and return if time permits.

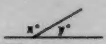
Quantitative Comparison Questions

Quantitative comparison questions emphasize the concepts of equalities, inequalities, and estimation. They generally involve less reading, take less time to answer, and require less computation than regular multiple-choice questions.

Directions: Each of the following questions consists of two quantities, one in Column A and one in Column B. You are to compare the two quantities and on the answer sheet blacken space

A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

Notes: **1. In certain questions, information concerning one or both of the quantities to be compared is centered above the two columns.**
2. In a given question, a symbol that appears in both columns represents the same thing in Column A as it does in Column B.
3. Letters such as x, n, and k stand for real numbers.

EXAMPLES	Column A	Column B	Answers	
E1.	2 × 6	2 + 6	● Ⓑ Ⓒ Ⓓ	(The answer is A because 12 is greater than 8.)
E2.	 180 − x	y	Ⓐ Ⓑ ● Ⓓ	(The answer is C because x + y = 180, thereby making 180 − x equal to y.)
E3.	p − q	q − p	Ⓐ Ⓑ Ⓒ ●	(The answer is D because nothing is known about either p or q.)

Quantitative comparison questions may not be as familiar to you as other types of questions. Therefore, give special attention to the directions ahead of time. To solve a quantitative comparison problem, you compare the quantities in the two columns and decide whether one quantity is greater than the other, whether the two quantities are equal, or whether the relationship cannot be determined from the information given. Remember that your answer should be:

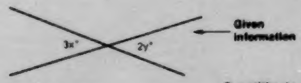
A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

Problems are clearly separated and the *quantities to be compared are always on the same line as the number of the problem (see example 2 below).* Figures and additional information provided for some problems appear *above* the quantities to be compared. The following are some practice problems with explanations to help you understand this type of question.

	Column A	Column B
1.	$(37)(\frac{1}{43})(58)$	$(59)(\frac{1}{43})(37)$

Because the numbers in this problem are fairly large, it may save time to study the multipliers first before attempting the calculations. Note that (37) and $(\frac{1}{43})$ appear in both quantities; thus, the only numbers left for you to compare are 58 and 59. Since $59 > 58$, the quantity on the right is greater and the correct answer is (B).

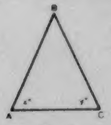
Column A **Column B**



	Column A	Column B
2.	x	y ← Quantities to be compared

To solve this problem, you need to recall that when two lines intersect, the vertical angles are equal. Therefore, $3x = 2y$ or $x = \frac{2}{3}y$, which implies that y is greater than x. The correct answer is (B). Note that the question asks for a comparison of x and y, not 3x and 2y.

Column A **Column B**



AB = BC

	Column A	Column B
3.	x	y

Since AB = BC, the angles opposite AB and BC are equal and, therefore, $x = y$. The correct answer is (C).

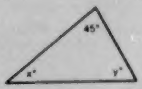
	Column A	Column B
4.	$x + 1$	$2x + 1$

Because both expressions contain a "1," the problem becomes one of comparing x with 2x. When you compare algebraic expressions, a useful technique is to consider zero and negative numbers for possible values of the unknown.

$2x > x$ for positive numbers
$2x = x$ for zero
$2x < x$ for negative numbers

The correct answer is (D), as the relationship cannot be determined from the information given. If you had been given that x was positive (that is, $x > 0$), the correct answer would have been B because 2x would be greater than x.

Column A **Column B**



Note: Figure is not drawn to scale.

$x < 45 < y$

	Column A	Column B
5.	y	90

Because the sum of the angles of a triangle is 180, $x + y + 45 = 180$ or $x + y = 135$. Since $x < 45$, it follows that $y > 90$. The answer is (A). In this problem you should not try to determine the answer from the appearance of the figure because the note indicates it is not drawn to scale.

Column A **Column B**

$x \neq 1$

	Column A	Column B
6.	$\frac{x^2 - 1}{x - 1}$	x

The condition $x \neq 1$ (read x is not equal to 1) is given because the algebraic fraction in column A is not defined for $x = 1$ (the denominator becomes zero). The solution of this problem involves simplifying the fraction in Column A as follows:

$$\frac{x^2 - 1}{x - 1} = \frac{(x + 1)(x - 1)}{x - 1} = x + 1$$

Therefore, the quantity in column A is equal to $x + 1$. Since $x + 1$ is always greater than x, the answer is (A).

	Column A	Column B
7.	Area of a triangle with altitude 4	Area of a triangle with base 5

To answer this question, you need to know how to find the area of a triangle. To find the area of a triangle, you need to know the length of a base and the altitude to that base. You can't find the "area of a triangle with altitude 4" without knowing the base, so the area of such a triangle could be any number depending on the length of the base. Likewise, you can't find the "area of a triangle with base 5" without knowing the length of the altitude. Since you can't tell anything about the two areas, the correct answer is (D).

THE TEST OF STANDARD WRITTEN ENGLISH

The questions on the Test of Standard Written English evaluate your ability to recognize standard written English, the language of most college textbooks and the English you will probably be expected to use in the papers you write for most college courses. The TSWE tests some basic principles of grammar and usage such as agreement of subject and verb and of pronoun and antecedent, and it also deals with more complicated writing problems such as whether or not the comparisons made in a sentence are logical.

No question asks you to define grammatical terms or identify particular verb tenses or correct spelling or capitalization. In only a few questions are punctuation marks like the semicolon or the apostrophe important in arriving at an answer.

The best preparation for the TSWE is to read widely in works by skilled writers and, of course, to get regular practice in both analyzing good models of prose and in writing and rewriting your own prose. However, you should also find it helpful to read the materials in this publication and become familiar with the directions and the two types of multiple-choice questions used on the test.

The test begins with 25 usage questions, then has 15 sentence correction questions, and ends with 10 more usage questions.

Usage Questions

Directions: The following sentences contain problems in grammar, usage, diction (choice of words), and idiom. Some sentences are correct. No sentence contains more than one error.

You will find that the error, if there is one, is underlined and lettered. Assume that all other elements of the sentence are correct and cannot be changed. In choosing answers, follow the requirements of standard written English.

If there is an error, select the one underlined part that must be changed to make the sentence correct, and blacken the corresponding space on your answer sheet. If there is no error, blacken answer space E.

EXAMPLES:

I. He spoke $\underset{A}{\underline{\text{bluntly}}}$ and $\underset{B}{\underline{\text{angrily}}}$ to $\underset{C}{\underline{\text{we}}}$ $\underset{D}{\underline{\text{spectators.}}}$ $\underset{E}{\underline{\text{No error}}}$

Ⓐ Ⓑ ● Ⓓ Ⓔ

II. She $\underset{A}{\underline{\text{works}}}$ $\underset{B}{\underline{\text{every day}}}$ $\underset{C}{\underline{\text{so that she}}}$ would become $\underset{D}{\underline{\text{financially}}}$ independent in her old age. $\underset{E}{\underline{\text{No error}}}$

● Ⓑ Ⓒ Ⓓ Ⓔ

Usage questions ask you to recognize writing that does not follow the conventions of standard written English. For example, they test the use of appropriate verb tenses, the agreement of pronoun with antecedent, the maintenance of parallel structure, and the use of idiomatic expressions. Each question consists of a sentence with four words or phrases underlined and lettered. You are to choose the one underlined part that must be changed to make the sentence acceptable in standard written English. Remember that the error must be one that you can correct by changing the underlined part. If you think that the sentence is all right as it is, mark the fifth choice, "No error." Read the entire sentence and consider each choice before you select your answer. Some of the sentences are correct, so don't hesitate to mark the "No error" choice if you can't find a mistake. Keep in mind, however, that errors of word choice as well as grammar or usage can be included.

Below are some sample usage questions and explanations of the correct answers.

1. **Most people will $\underset{A}{\underline{\text{find it easier}}}$ to adjust to the metric system $\underset{B}{\underline{\text{if you become}}}$ familiar $\underset{C}{\underline{\text{with}}}$ it before the changeover $\underset{D}{\underline{\text{occurs.}}}$ $\underset{E}{\underline{\text{No error}}}$**

   The problem in this sentence is that of shift in person. The writer begins talking about "most people" and then changes to "you." The sentence could be corrected by changing either "most people" or "you." However, "you" is underlined and "most people" is not, so, according to the directions, "you" is the only part of the sentence where the change can be made. The construction in choice (A) is acceptable, the preposition in choice (C) is the accepted idiom, and the number and tense of the verb in (D) agree. The correct answer is (B). The revised sentence would read "Most people will find it easier. . . if they become. . ."

2. **Many travelers claim $\underset{A}{\underline{\text{having seen}}}$ the Abominable Snowman, $\underset{B}{\underline{\text{but}}}$ no one has proved that $\underset{C}{\underline{\text{such a creature}}}$ $\underset{D}{\underline{\text{actually}}}$ exists. $\underset{E}{\underline{\text{No error}}}$**

The answer is (A). In the context of this sentence, the "having seen" is not acceptable. In standard written English, you should write "Many travelers claim to have seen. . ." Choice (B), "but," links the two parts of the sentence in a logical and appropriate way, and the expression at (C) and the adverb at (D) are also acceptable.

3. **One of the activities of women's organizations**
A
is to encourage projects that will make life easier for
B C D
working mothers. No error
E

Although the ideas expressed in the sentence might be conveyed with different words and expressions, the sentence contains no grammatical, idiomatic, logical, or structural errors. The answer is (E).

4. **Holes in the skulls of people from certain ancient civilizations suggests that modern physicians are not the only**
A B
doctors who have attempted to operate on the brain.
C D
No error
E

The error is that of subject-verb agreement. Because the subject of the sentence, "Holes," is plural, the verb must also be plural; "suggests" should be changed to "suggest." Always check to see that the verb agrees with the subject, especially when the verb is separated from the subject by a phrase or a clause or when the verb comes before the subject. (B), (C), and (D) are all acceptable in standard written English. The answer is (A).

Sentence Correction Questions

Directions: In each of the following sentences, some part or all of the sentence is underlined. Under each sentence you will find five ways of phrasing the underlined part. Select the answer that produces the most effective sentence, one that is clear and exact, without awkwardness or ambiguity, and blacken the corresponding space on your answer sheet. In choosing answers, follow the requirements of standard written English, but do not make a choice that changes the meaning of the original sentence.

Answer (A) is always the same as the underlined part. Choose answer (A) if you think the original sentence needs no revision.

EXAMPLE:

Ms. Rose planning to teach a course in biology next summer.

(A) planning (B) are planning
(C) have planned
(D) with a plan (E) plans

Ⓐ Ⓑ Ⓒ Ⓓ ●

Many of these questions present complex faults in the logic or structure of a sentence, errors that require substantial rewriting. The directions ask you to select the *best* version of the sentence, so you must consider *effectiveness* of expression as well as correctness. Look at each of the five choices carefully before you select your answer. If you think more than one answer is acceptable, read each choice as part of the entire sentence; sometimes a choice makes sense by itself but not as part of the sentence. Some of the choices may be grammatically correct and logically and structurally consistent with the rest of the sentence, but they may still not be the correct response because they are wordy or awkward or because they lack force or emphasis.

Each of the choices is likely to deal with more than one writing problem. If you are not careful, you may pick an answer that corrects one problem but creates a new one. You are more likely to pick the correct response if you ask yourself questions like these: Is this sentence grammatically correct? Is it a complete sentence? Are the elements in the sentence parallel? Are the parts of the sentence linked in an appropriate manner? Are the ideas expressed or compared logically? Is the wording clear and concise?

Below are some sample sentence correction questions and explanations of the correct answers.

1. **The shorter bearpaw snowshoes are the best choice if you are looking for an easy to lift and maneuver model.**

(A) an easy to lift and maneuver model
(B) a model that is easy to lift and maneuver
(C) an easy model as far as lifting and maneuvering goes
(D) a model with ease of lifting and also maneuver
(E) an easily lifted and also maneuvered model

The sentence tests directness and effectiveness of expression. Choices (A) and (E) are grammatically correct, but are not as well expressed as choice (B). Choice (C) is awkward. Choice (D) does not maintain parallel structure. Only choice (B) is clear, concise, and idiomatic.

2. **Althea Gibson was the first Black American to win major tennis championships and played in the 1950s.**

(A) Althea Gibson was the first Black American to win major tennis championships and played in the 1950s.
(B) Althea Gibson, being the first Black American to win major tennis championships, and playing in the 1950s.
(C) Althea Gibson, playing in the 1950s, being the first Black American to win major tennis championships.
(D) Althea Gibson, who played in the 1950s, was the first Black American to win major tennis championships.
(E) Althea Gibson played in the 1950s, she was the first Black American to win major tennis championships.

When a sentence makes more than one statement, it is customary to indicate the relationship between these statements. If the statements are equal in importance, rank, or degree, the writer can use a conjunction like "and" or "but." This method of linking is called coordination. If one statement is to be emphasized more than the other, the writer can use a conjunction like "because" or "since" or other words like "who" or "that." This method of linking is called subordination. Sometimes, the intention of the writer determines which method is most appropriate. In the underlined sentence about Althea Gibson, the writer has used "and" to join two statements about the athlete. The sentence (and therefore choice A) is not effective, however, because "and" does not adequately convey the relationship between the two statements. A more interesting sentence would indicate which one of the statements was more important to the writer. The only choice that does so is (D). It is a well-constructed sentence, one with appropriate subordination and emphasis.

3. Placed in the time capsule, the scientists did so in the hope that the documents would be found by future generations.

(A) Placed in the time capsule, the scientists did so in the hope that the documents
(B) When they were placed in the time capsule, the scientists did it in the hope that the documents
(C) By placing them in the time capsule, the hope of the scientists was that the documents
(D) The scientists placed the documents in the time capsule in the hope that these papers
(E) The placing of the documents in the time capsule by the scientists was in the hope that these papers

The primary problem with the sentence is that it contains a dangling modifier. A dangling modifier is a phrase, a clause, or an adjective that seems to modify a word that it was not intended to modify. Dangling modifiers often result in sentences that are ambiguous or illogical. Modifiers should usually be placed next to or close to the word they are intended to modify to avoid problems of ambiguity. Because of the way the original sentence and choice (A) are constructed, the phrase "Placed in the time capsule" seems to modify "the scientists." However, this relationship makes no sense. The writer doesn't mean that the scientists were placed in the time capsule but that the scientists placed the documents in the time capsule. Choice (B) contains a similar error and adds to it by using the pronoun "it" without a clear antecedent. Choice (C) has a dangling modifier problem and is further weakened by the distance between "them" and the word it refers to, "documents." Choice (E) does not have a dangling modifier, but it uses the passive voice in an awkward manner and is unnecessarily wordy. The intended idea of the original sentence is expressed clearly and logically only in choice (D), the correct answer.

4. Being as it was a full moon, the tides were exceptionally high when the storm struck.

(A) Being as it was a full moon
(B) With the moon as full
(C) Due to there being a full moon
(D) Because the moon was full
(E) The moon was full

This question presents problems in wording and in structure. Choices (A), (B), and (C) use expressions that are not ordinarily used in standard written English ("Being as it was," "With the moon as," "Due to there being"). Choice (E) joins two sentences with a comma, a method that is not acceptable in standard written English.

What is needed is some indication of the relationship between the first part of the sentence and the second. Choice (D) solves the problem—the word "because" clarifies the relationship by indicating that the situation described in the first part affected the situation described in the second. The answer is choice (D).

THE SAMPLE TEST

The sample test that follows is the edition of the SAT given on April 4, 1981 (in New York State on June 6, 1981). It includes only five of the six sections in the actual test. The experimental section, Section 4 in this edition, has been omitted because it contains questions that may be used in future editions of the SAT and does not count toward the scores. The sample test will be most helpful if you take it under conditions as close as possible to those of the actual test:

- Set aside two-and-one-half hours when you will not be interrupted, so that you can complete all of the test in one sitting.
- Sit at a desk with no other papers or books. You can't take a calculator, dictionary, other books, or notes into the test room.
- Have a kitchen timer or clock in front of you for timing yourself on the sections.
- Allow yourself only 30 minutes for each section of the test.
- Fill in the sample answer sheet on page 47 just as you would at a regular test.
- Read the instructions below. They are reprinted from the back cover of the test book. When you take the test, you will be asked to read them before you begin answering questions.
- After you finish the practice test, read "How to Score Your Sample Test," on page 44.

SCHOLASTIC APTITUDE TEST

and Test of Standard Written English

You will have three hours to work on the questions in this test book, which is divided into six 30-minute sections. The supervisor will tell you when to begin and end each section. If you finish before time is called, you may check your work on that section, but you are not to work on any other section.

Do not worry if you are unable to finish a section or if there are some questions you cannot answer. Do not waste time puzzling over a question that seems too difficult for you. You should work as rapidly as you can without sacrificing accuracy.

Students often ask whether they should guess when they are uncertain about the answer to a question. Your test scores will be based on the number of questions you answer correctly minus a fraction of the number you answer incorrectly. Therefore, it is improbable that random or haphazard guessing will change your scores significantly. If you have some knowledge of a question, you may be able to eliminate one or more of the answer choices as wrong. It is generally to your advantage to guess which of the remaining choices is correct. Remember, however, not to spend too much time on any one question.

Mark all your answers on the separate answer sheet. Mark only one answer for each question. Since the answer sheet will be machine scored, be sure that each mark is dark and that it completely fills the answer space. In each section of the answer sheet, there are spaces to answer 50 questions. When there are fewer than 50 questions in a section of your test, mark only the spaces that correspond to the question numbers. Do not make stray marks on the answer sheet. If you erase, do so completely, because an incomplete erasure may be scored as an intended response.

You may use the test book for scratchwork, but you will not receive credit for information written there.

DO NOT OPEN THIS BOOK UNTIL THE SUPERVISOR TELLS YOU TO DO SO.

1

Sample Test

SECTION I

Time—30 minutes

45 QUESTIONS

For each question in this section, choose the best answer and blacken the corresponding space on the answer sheet.

Each question below consists of a word in capital letters, followed by five lettered words or phrases. Choose the word or phrase that is most nearly opposite in meaning to the word in capital letters. Since some of the questions require you to distinguish fine shades of meaning, consider all the choices before deciding which is best.

Example:

> GOOD: (A) sour (B) bad (C) red
> (D) hot (E) ugly
> ○ ● ○ ○ ○

1. BAN: (A) borrow (B) regret
(C) permit (D) conquer (E) exaggerate

2. COMPRESSION: (A) equality (B) expansion
(C) exposure (D) endurance (E) excitement

3. FRAUDULENT: (A) dynamic (B) masterly
(C) possible (D) genuine (E) abundant

4. PARASITE: (A) expert (B) imposter
(C) instigator (D) self-assured snob
(E) self-sufficient individual

5. SPARSE: (A) thick (B) tidy
(C) wealthy (D) round (E) sticky

6. DENOUNCE: (A) overstate (B) acclaim
(C) destroy (D) refuse (E) hasten

7. FLY-BY-NIGHT: (A) unbalanced (B) moderate
(C) permanent (D) incredible (E) modern

8. STERILE: (A) venal (B) productive
(C) generous (D) variegated (E) unalloyed

9. REJOICE: (A) defend against (B) shrug off
(C) criticize (D) bemoan (E) discriminate

10. FANATICISM: (A) optimism (B) hedonism
(C) penitence (D) didacticism (E) apathy

11. VESTIGIAL: (A) fully developed
(B) publicly announced (C) offensive
(D) provincial (E) miraculous

12. DIVERT: (A) bore (B) rescue (C) espouse
(D) judge fairly (E) question relentlessly

13. HACKNEYED: (A) integrated (B) appealing
(C) inventive (D) acceptable (E) improper

14. ALACRITY: (A) indolence (B) bravery
(C) wisdom (D) pungency (E) retention

15. BURGEON: (A) fail to prove
(B) shrivel and die (C) pursue and capture
(D) disobey regulations (E) speak incoherently

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five lettered words or sets of words. Choose the word or set of words that best fits the meaning of the sentence as a whole.

Example:

> Although its publicity has been —, the film itself is intelligent, well-acted, handsomely produced, and altogether —.
>
> (A) tasteless. .respectable (B) extensive. .moderate
> (C) sophisticated. .amateur (D) risqué. .crude
> (E) perfect. .spectacular
> ● ○ ○ ○ ○

16. Rather than — wagon trains and — the pioneer's movement westward, many American Indians acted as guides and companions.
(A) encountering. .helping
(B) seeking. .encouraging
(C) attacking. .hindering
(D) welcoming. .allowing
(E) repulsing. .following

17. Although Ricardo looked tired, he was enormously — by Maria's election, a victory they had both labored so hard for.
(A) elated (B) baffled (C) fatigued
(D) surprised (E) exasperated

18. The earth's atmosphere — surface temperatures by distributing heat absorbed on the sunlit side to the side in shadow.
(A) warms (B) resists (C) generates
(D) moderates (E) amasses

19. Professor Greene's —, detailed, and lucid lectures match the gravity of her subject.
(A) blithe (B) studious (C) jaundiced
(D) obscure (E) uproarious

20. It would seem that the — of science, the building and making of things, have stirred the American imagination, but that the — roots of such activities have not.
(A) applications. .theoretical
(B) hypotheses. .practical
(C) formulas. .actual
(D) offshoots. .extrinsic
(E) products. .future

GO ON TO THE NEXT PAGE

Copyright © 1981 by Educational Testing Service. All rights reserved.

Each passage below is followed by questions based on its content. Answer all questions following a passage on the basis of what is stated or implied in that passage.

(The passages for this test have been adapted from published material. The ideas contained in them do not necessarily represent the opinions of the College Board or Educational Testing Service.)

(This passage was written in 1961.)

The classics are the great works of human intelligence that, whether produced in Israel, Greece, Rome, or England, have demonstrated the capacity for being understood over a long period of time and in a variety of places. They introduce the student to the significant ideas common to Western civilization. But most students are so busy learning the language of the classics that they never become acquainted with the ideas in them.

I know not how many Gauls Caesar slaughtered. But I know that learning Latin from the *Gallic Wars* has killed the interest of generations of Americans in the classics. More than sixty years ago, a committee of educational consultants affirmed that the *Gallic Wars* did not belong in the classroom. According to the committee, "The book is altogether too difficult for beginners; it is too military in content to be generally interesting. Its vocabulary is too restricted to marches, sieges, and battles to afford the best introduction to subsequent reading." Yet, of the glory that was Greece and Rome, this is what high school students first encounter. And since they are forced to further waste time by reading it in the original, this is often their only encounter.

The stubborn attachment to Caesar springs from an underlying insistence that the ultimate purpose of teaching the classics is to teach Latin and that therefore the simple style of Caesar is appropriate. For the same reason, *Beowulf* has been set as an impenetrable barrier to a love of English literature by those who believe that the study of English literature must begin with the knowledge of the Anglo-Saxon language.

However, if we use the classics to offer a glimpse of the seminal ideas of Western culture, quite different texts can be held forth. All students can read some Aristophanes or Thucydides or Chaucer with interest. Inclusive lists are of no importance, for these texts should be the beginning, rather than the sum, of their reading. And they should begin by reading the good English translations that are available. No translation is, of course, the equivalent of the original. But the question is not whether *The Clouds* should be read in Greek or in English; the question is whether it should be read in English or not read at all. It is more likely that a few students will be drawn to study the language through knowledge of its literature than that any will be drawn to the literature through drill in the language.

21. Which of the following best expresses the main idea of the passage?
    - (A) The *Gallic Wars* is more appropriate today than it was many years ago.
    - (B) Most students are more interested in learning languages than in learning history.
    - (C) The place of the classics in high school has been greatly overrated.
    - (D) Good English translations of the classics are usually stylistically superior to the originals.
    - (E) The classics should be used in the high school to present meaningful ideas.

22. It can be inferred that the committee of educational consultants mentioned in the passage recommended that
    - (A) Latin be studied in the high schools, but from a text other than Caesar's *Gallic Wars*
    - (B) Caesar's *Gallic Wars* continue to be studied in American schools, but in a good English translation
    - (C) the study of the classics be dropped from American school curricula
    - (D) Latin be studied from Caesar's *Gallic Wars* because the style, if not the content, is appropriate for beginners
    - (E) an abridged version of the *Gallic Wars*, omitting much of the military content, be published

23. Which of the following would the author probably recommend as the best approach to the study of classics in American high schools?
    - (A) The classics should not be studied in any form.
    - (B) The classics should be studied in the original language, but a restricted vocabulary should be used.
    - (C) An appreciation of the classics should grow out of a thorough knowledge of them in their original languages.
    - (D) The classics should be read in English translations.
    - (E) The classics should be studied in relation to the literary theories of the time in which each work was written.

24. The author refers to Aristophanes and Thucydides as examples of classical authors who are
    - (A) included in most high school courses
    - (B) of intrinsic interest to young people
    - (C) destroyed by awkward translations
    - (D) of no interest because of their difficulty
    - (E) difficult to relate to the rest of the literature curriculum

25. According to the author, the use of the classics in high schools was primarily based on the instructor's desire to
    - (A) stimulate students' interest in languages
    - (B) teach students self-discipline and promote a methodical approach to learning
    - (C) develop in students a familiarity with a particular language
    - (D) give young people an appreciation of contemporary achievements
    - (E) familiarize students with the important names in Western literature

GO ON TO THE NEXT PAGE

The renewed interest in ragtime is one of the most curious cases of changing musical taste in recent years. The resurgence began in 1970 when a recording of eight rags by the inimitable Scott Joplin appeared.

This recording was not a rediscovery of forgotten material, for ragtime had remained popular with jazz and Dixieland audiences. Jazz musicians had altered the steady beat of the music, adding syncopations and, in the process, eliminating the contrasts that highlighted the melodies; Dixieland musicians had done the pieces very fast, damaging the lyrical qualities. But in the 1970 recording, the music was played exactly as Joplin had written it. What emerged was a delicate music, with graceful melodies and compelling rhythms, that captivated a large new audience. Since then, modified ragtime recordings have inundated the market.

Ragtime was created during the last years of the nineteenth century by itinerant black musicians of the Midwest out of previously disparate elements. From the minstrel song they took the melodic style characteristic of banjo accompaniment; from the march they took the bass figures and rhythms; from the cakewalk they took a mock-serious combination of stately procession and energetic improvisation.

Present-day ragtime enthusiasts seem to delight in emphasizing the difficulties that faced ragtime composers. However, to cast a man such as Joplin as an "all-but-forgotten black American genius" not only ignores important elements of Joplin's life, but also misrepresents the development of popular music in the twentieth century. Within months of the publication of "Maple Leaf Rag," Joplin became a rich and famous man and remained so until his death. Also, the relative decline of ragtime cannot be blamed on hostility to black culture, since another form of black music, jazz, became the new rage when the original ragtime boom receded.

The attempt to assimilate ragtime into the romantic myth of the artist neglected by society can be seen as part of an effort to recast ragtime as "serious" or "classical" music. Admittedly, of all the forms of popular music, ragtime is the most amenable to "classical" treatment. Unlike jazz and folk music, ragtime is not improvisatory. It is fully written out, with no reliance on aural tradition. And unlike the work of the popular songwriters, ragtime is written for the solo piano—the preeminent classical instrument. Nevertheless, it is also true that the rag is a modest form, with an elementary structure and technique. Burdening ragtime with an unnecessary load of musical pretension and political meaning may sink this music into a deeper oblivion than the relative neglect from which it has been rescued.

26. According to the passage, Dixieland musicians altered the original style of ragtime music by

(A) highlighting the melodies
(B) adding syncopations
(C) playing the music very fast
(D) giving the music a steady beat
(E) omitting the lyrics of the music

27. By the statement "previously disparate elements" (lines 22-23), the author is referring to the fact that

(A) blacks had never before organized to develop an authentic style of music
(B) the source of a particular style of music had never before been attributed to a particular region of the country
(C) ragtime was the first black music to attract a large and varied audience
(D) ragtime was the first black music in which the piano rather than the banjo was the dominant instrument
(E) ragtime was created by mixing various aspects of other types of music

28. It can be inferred from the passage that, when it was written, Scott Joplin's music was

(A) overrated by critics
(B) dismissed as unprofitable
(C) ignored by audiences
(D) acknowledged as unique and praiseworthy
(E) considered classical rather than popular

29. With which of the following statements regarding ragtime would the author probably agree?

I. What is best in ragtime may be lost in the effort to make it something it is not.
II. Joplin's ragtime did more than any other style of music to revolutionize American music.
III. Ragtime differs in some important ways from most other popular music.

(A) I only (B) I and II only (C) I and III only
(D) II and III only (E) I, II, and III

30. In developing the passage, the author does all of the following EXCEPT

(A) present evidence from other people to support the views expressed
(B) give historical information regarding the origins of ragtime music
(C) attempt to convey a feeling of the nature of ragtime music
(D) offer statements to refute certain aspects of the opposing point of view
(E) describe a potential consequence of the opposing point of view

GO ON TO THE NEXT PAGE

Select the word or set of words that best completes each of the following sentences.

31. Many of the officers were unsympathetic, even —, to their new and inexperienced commander.

(A) lukewarm (B) indulgent
(C) servile (D) hostile (E) impartial

32. The rural Republican and the urban Democrat were immensely different, and the — that separated them was not — successfully during the postwar era.

(A) bulwark. .built
(B) goal. .achieved
(C) gulf. .bridged
(D) economy. .widened
(E) electorate. .enfranchised

33. Whereas his predecessors were careful to stay within the boundaries of factual evidence, Dr. Artman delivered an address that was — by pretentious and — theorizing.

(A) marred. .irresponsible
(B) undermined. .unassuming
(C) supported. .flawless
(D) refuted. .unverifiable
(E) represented. .unquestionable

34. Because of their extremely simple structures, viruses have proved invaluable to those scientists who are interested in constructing — for the larger, more complex molecules.

(A) a cure (B) a response (C) an edifice
(D) a universe (E) a paradigm

35. Many early Americans, for all their dislike of monarchy, firmly believed in — society.

(A) a democratic (B) an idealistic
(C) a stratified (D) a historic (E) a flexible

Each question below consists of a related pair of words or phrases, followed by five lettered pairs of words or phrases. Select the lettered pair that best expresses a relationship similar to that expressed in the original pair.

Example:

YAWN : BOREDOM :: (A) dream : sleep
(B) anger : madness (C) smile : amusement
(D) face : expression (E) impatience : rebellion
Ⓐ Ⓑ ● Ⓓ Ⓔ

36. FOOD : STARVATION :: (A) liquor : inebriation
(B) water : saturation (C) heat : inflammation
(D) privacy : isolation (E) air : suffocation

37. TAPE RECORDER : EAR :: (A) radio : antenna
(B) camera : eye (C) phonograph : volume
(D) journal : hand (E) telephone : speech

38. SUMMONS : ATTENDANCE ::
(A) allowance : money (B) bill : payment
(C) purchase : article (D) question : examination
(E) continuation : action

39. HOPE : DESPAIRING :: (A) confidence : friendly
(B) respect : governing (C) wittiness : humorous
(D) jollity : gloomy (E) unconcern : poised

40. ACT : TRAGEDY :: (A) stanza : poem
(B) palette : artist (C) inventor : machine
(D) gate : fence (E) cover : book

41. ENIGMA : MYSTERIOUS ::
(A) beginning : vague
(B) decision : cautious
(C) memory : forgotten
(D) meeting place : random
(E) turning point : significant

42. WHELP : DOG :: (A) mule : horse
(B) kitten : pet (C) child : human being
(D) criminal : community (E) school : fish

43. ANNOY : ABUSE :: (A) alter : change
(B) encourage : prod (C) clamor : silence
(D) violate : revolt (E) consider : discover

44. PARTICLE : AGGREGATE :: (A) athlete : rivalry
(B) molecule : surface (C) asteroid : planet
(D) animal : herd (E) motor : power

45. UNWITTING : INTENTION ::
(A) inconspicuous : strategy
(B) alone : assistance
(C) lavish : expense
(D) impractical : explanation
(E) futile : disappointment

STOP

IF YOU FINISH BEFORE TIME IS CALLED, YOU MAY CHECK YOUR WORK ON THIS SECTION ONLY.
DO NOT WORK ON ANY OTHER SECTION IN THE TEST.

2

SECTION II

Time—30 minutes

25 QUESTIONS

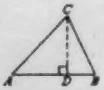
In this section solve each problem, using any available space on the page for scratchwork. Then decide which is the best of the choices given and blacken the corresponding space on the answer sheet.

The following information is for your reference in solving some of the problems.

Circle of radius r: Area $= \pi r^2$; Circumference $= 2\pi r$
The number of degrees of arc in a circle is 360.
The measure in degrees of a straight angle is 180.

Definitions of symbols:

$=$	is equal to	$\leq$	is less than or equal to
$\neq$	is unequal to	$\geq$	is greater than or equal to
$<$	is less than	$\parallel$	is parallel to
$>$	is greater than	$\perp$	is perpendicular to



Triangle: The sum of the measures in degrees of the angles of a triangle is 180.

If $\angle CDA$ is a right angle, then

(1) area of $\triangle ABC = \dfrac{AB \times CD}{2}$

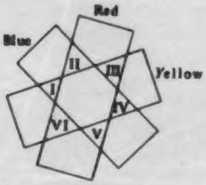
(2) $AC^2 = AD^2 + DC^2$

Note: Figures which accompany problems in this test are intended to provide information useful in solving the problems. They are drawn as accurately as possible EXCEPT when it is stated in a specific problem that its figure is not drawn to scale. All figures lie in a plane unless otherwise indicated. All numbers used are real numbers.

---

1. If $x^2 - 1 = y$ and $x = 3$, then $y^2 =$

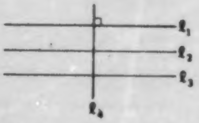
(A) 81 (B) 64 (C) 9 (D) 8 (E) 4



Red + Blue = Purple
Red + Yellow = Orange
Blue + Yellow = Green

2. The figure above shows strips of colored glass that overlap to form other colors as shown by the color chart. Which two labeled triangular regions would be green?

(A) I and III
(B) I and IV
(C) II and V
(D) III and VI
(E) IV and VI



3. In the figure above, if $\ell_1 \parallel \ell_2$, $\ell_2 \parallel \ell_3$, and $\ell_1 \perp \ell_4$, which of the following statements must be true?

I. $\ell_1 \parallel \ell_3$
II. $\ell_2 \perp \ell_4$
III. $\ell_3 \perp \ell_4$

(A) None (B) I only (C) I and II only
(D) II and III only (E) I, II, and III

4. If $100 \leq k \leq 400$ and k is a multiple of 5, 6, 7, and 10, then $k =$

(A) 105
(B) 150
(C) 210
(D) 300
(E) 350

GO ON TO THE NEXT PAGE

5. In a restaurant where the sales tax on a \$4.00 lunch is \$0.24, what will be the sales tax due on a \$15.00 dinner?

(A) \$0.60
(B) \$0.75
(C) \$0.90
(D) \$1.20
(E) \$1.74

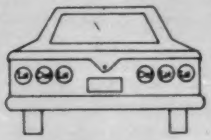
1st Row	0 1 0
2nd Row	0 1 1 0
3rd Row	0 1 2 1 0
4th Row	0 1 3 3 1 0
5th Row	0 1 4 6 4 1 0
6th Row —	

6. The nonzero numbers above form a triangular array. Beginning with the second row, each nonzero number in a row is the sum of the two numbers nearest to it in the row immediately above. If a sixth row is added in this fashion, what will be the sum of all the numbers in the sixth row?

(A) 8 (B) 10 (C) 16 (D) 32 (E) 64

7. If x is a positive integer and $x^2 + x = n$, which of the following could be the value of n?

(A) 14
(B) 15
(C) 18
(D) 23
(E) 30



8. If one more of the lit signals on the rear of the car above were out, what per cent of all the rear signals would then be lit?

(A) 25% (B) $33\frac{1}{3}$% (C) 50%

(D) $66\frac{2}{3}$% (E) 75%

9. If $111{,}111 + N = 181{,}111$, then $N =$

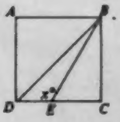
(A) 7×10^3
(B) 7×10^4
(C) 7×10^5
(D) 8×10^4
(E) 9×10^4

10. In the United States in a certain year, food production per person was 15 per cent greater than food consumption per person. If the average daily consumption per person in the United States in that year was 3,000 calories, what was the average daily production (in calories) per person in that year?

(A) 3,200 (B) 3,450 (C) 3,600
(D) 3,850 (E) 4,500

11. 0.06 is the ratio of 6 to

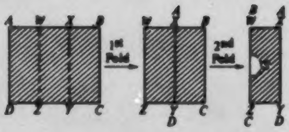
(A) 1,000 (B) 100 (C) 10
(D) $\frac{1}{10}$ (E) $\frac{1}{100}$



12. In the figure above, $ABCD$ is a square. If $BE = 2EC$, then $x =$

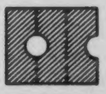
(A) 100 (B) 110 (C) 120
(D) 150 (E) 160

GO ON TO THE NEXT PAGE

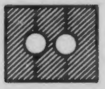


13. In the figure above, a rectangular piece of paper *ABCD* is folded along dotted line *WZ* so that *A* is on top of *X* and *D* is on top of *Y* and then folded along *XY* so that *B* is on top of *W* and *C* is on top of *Z*. A small semicircle *S* with diameter on *BC* is cut out of the folded paper. If the paper is unfolded, which of the following could be the result?

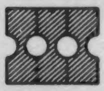
(A)



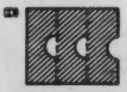
(B)



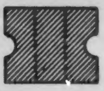
(C)



(D)



(E)



---

14. If $5x - 3y = 8$ and $x = \frac{7y}{5}$, then $y =$

(A) $\frac{5}{4}$

(B) $\frac{8}{5}$

(C) 2

(D) $\frac{8}{3}$

(E) 4

15. Amy is twice as old as Bill. Five years ago she was 3 times as old as Bill was then. How old is Bill now?

(A) 20
(B) 15
(C) 10
(D) 5
(E) It cannot be determined from the information given.

16. $(3x^2 - 4x + 7) - (2x + 1)(x - 5) =$

(A) $x^2 + 5x + 12$

(B) $x^2 - 5x + 12$

(C) $x^2 - 5x + 2$

(D) $5x^2 - 13x + 2$

(E) $5x^2 + 13x + 12$

17. In $\triangle ABC$, the ratio of the length of side AB to the perimeter is 1 to 3. What is the ratio of the length of side BC to the perimeter?

(A) $\frac{1}{4}$

(B) $\frac{1}{3}$

(C) $\frac{5}{12}$

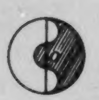
(D) $\frac{1}{2}$

(E) It cannot be determined from the information given.

18. $\frac{(N-2)(N-4)(N-6)(N-8)-1}{2}$ is an integer if N is equal to

(A) 1 only
(B) 2 only
(C) 9 only
(D) any odd integer
(E) any even integer

GO ON TO THE NEXT PAGE



19. Three equal semicircles are drawn on a diameter of the circle with center O as shown above. If the area of circle O is 9π, then the area of the shaded region is

(A) $\frac{7\pi}{2}$ (B) 4π (C) $\frac{9\pi}{2}$ (D) 5π (E) $\frac{11\pi}{2}$

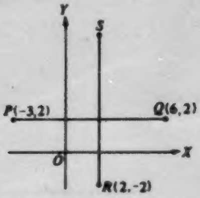
---

20. If each angle of quadrilateral $ABCD$ measures less than 180° and if 3 of its angles each measure x°, which of the following must be true?

(A) $x > 60$
(B) $x = 60$
(C) $x < 60$
(D) $ABCD$ is a parallelogram but not a square.
(E) $ABCD$ is a square.

21. The set P consists of all numbers which are the sum of 3 consecutive prime numbers. For example, the number 109 is in P, since $31 + 37 + 41 = 109$. The least prime number in P is

(A) 13
(B) 17
(C) 19
(D) 23
(E) 31



22. The coordinates of points P, Q, and R are shown in the figure above. If $PQ = RS$ and $PQ \perp RS$, what are the coordinates of S?

(A) (2, 6) (B) (2, 7) (C) (2, 8)
(D) (2, 9) (E) (7, 2)

23. How many minutes will it take a rocket to travel 4,000 miles if its average rate is 100 miles every t seconds?

(A) $\frac{2t}{3}$

(B) $\frac{3t}{2}$

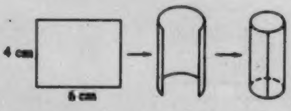
(C) $\frac{2}{3t}$

(D) $40t$

(E) $2{,}400t$

24. If the average of x, y, and 80 is 6 more than the average of y, z, and 80, what is the value of $x - z$?

(A) 2
(B) 3
(C) 6
(D) 18
(E) It cannot be determined from the information given.



25. The figure above shows how a rectangular piece of paper is rolled to form a cylindrical tube. If it is assumed that the 4-centimeter sides of the rectangle meet with no overlap, what is the area, in square centimeters, of the circular base region enclosed?

(A) 16π

(B) 9π

(C) 4π

(D) $\frac{9}{\pi}$

(E) $\frac{4}{\pi}$

S T O P

IF YOU FINISH BEFORE TIME IS CALLED, YOU MAY CHECK YOUR WORK ON THIS SECTION ONLY.
DO NOT WORK ON ANY OTHER SECTION IN THE TEST.

SECTION III

Time—30 minutes

40 QUESTIONS

For each question in this section, choose the best answer and blacken the corresponding space on the answer sheet.

Each question below consists of a word in capital letters, followed by five lettered words or phrases. Choose the word or phrase that is most nearly opposite in meaning to the word in capital letters. Since some of the questions require you to distinguish fine shades of meaning, consider all the choices before deciding which is best.

Example:

GOOD: (A) sour (B) bad (C) red
(D) hot (E) ugly
○ ● ○ ○ ○

1. ORDINARY: (A) numerical (B) rational
(C) impolite (D) staunch (E) abnormal

2. ATHEIST: (A) believer (B) scholar
(C) recluse (D) expatriate (E) pauper

3. FLICKER: (A) rise slowly (B) burn steadily
(C) warm completely (D) fume (E) collide

4. SERRATED: (A) undervalued (B) aggressive
(C) smooth and even (D) loose and flexible
(E) supremely confident

5. COSMOPOLITAN: (A) indecisive (B) ineffectual
(C) antagonistic (D) parochial (E) deferential

6. HYPOCRITICAL: (A) guileless (B) eccentric
(C) perspicacious (D) untrustworthy
(E) sagacious

7. ACCOLADE: (A) appetizer (B) censure
(C) recoil (D) seizure (E) referendum

8. PROPOUND:
(A) remove from consideration
(B) defend without evidence
(C) avoid without reason
(D) view with humility
(E) rescue from imprisonment

9. EBULLIENT: (A) staid (B) anxious
(C) feminine (D) unique (E) respectful

10. EXCULPATE: (A) require (B) yield
(C) reinstate (D) vivify (E) convict

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five lettered words or sets of words. Choose the word or set of words that best fits the meaning of the sentence as a whole.

Example:

Although its publicity has been —, the film itself is intelligent, well-acted, handsomely produced, and altogether —.

(A) tasteless. .respectable (B) extensive. .moderate
(C) sophisticated. .amateur (D) risqué. .crude
(E) perfect. .spectacular
● ○ ○ ○ ○

11. As tourism and industry develop along the shores of the Gulf of California, it is possible that the increased human — may drive the gray whales from their — calving sites.
(A) access. .forsaken
(B) knowledge. .impending
(C) activity. .preferred
(D) indifference. .destined
(E) concern. .despoiled

12. Because of its immense scope, critics have rightly referred to the book as the most — examination of minority concerns in the United States ever written.
(A) deficient (B) intuitive (C) obscure
(D) unaspiring (E) comprehensive

13. The — of personality that strikes us in Julia Cameron's best pictures gives her work a coherence that is — in the work of less accomplished personal photographers.
(A) idea. .compounded (B) image. .implied
(C) theory. .evolving (D) force. .lacking
(E) feeling. .restrained

GO ON TO THE NEXT PAGE

14. In addition to the traditional functions of collecting and displaying, a museum establishes the official — of an artist's works, — them a prestige unattainable in a commercial showcase.

(A) utility. .proposing for
(B) subjectivity. .sharing with
(C) style .conceding to
(D) importance. .conferring on
(E) destiny. .diverting from

15. Although Spalding — the importance of the physical necessities of life, her most successful endeavor was the — of the condition of the impoverished.

(A) deprecated. .alleviation
(B) emphasized. .investigation
(C) accentuated. .amelioration
(D) epitomized. .delineation
(E) disregarded. .desecration

Each question below consists of a related pair of words or phrases, followed by five lettered pairs of words or phrases. Select the lettered pair that best expresses a relationship similar to that expressed in the original pair.

Example:

YAWN : BOREDOM :: (A) dream : sleep
(B) anger : madness (C) smile : amusement
(D) face : expression (E) impatience : rebellion
Ⓐ Ⓑ ● Ⓓ Ⓔ

16. SANDBOX : PLAY ::
(A) picture : see (B) office : work
(C) library : publish (D) restaurant : guide
(E) kindergarten : manipulate

17. TENANT : RENT :: (A) salesperson : commission
(B) performer : ticket (C) investor : interest
(D) client : fee (E) professor : tuition

18. DAWN:DAY :: (A) moon:night (B) star:sun
(C) week:year (D) birth:life (E) beginning:end

19. MALLET:POLO :: (A) putter:club
(B) bat:baseball (C) field:football
(D) puck:hockey (E) lane:bowling

20. TERRESTRIAL:LUNG :: (A) marsupial:pouch
(B) floral:root (C) aquatic:gill
(D) perennial:seed (E) canine:mouth

21. DEFAME:REPUTATION ::
(A) demoralize:misery (B) challenge:opinion
(C) promote:talent (D) disfigure:appearance
(E) jeopardize:peril

22. NEUTER:GENDER :: (A) sincere:truth
(B) uniform:errors (C) futile:hindrance
(D) nimble:energy (E) destitute:possessions

23. HONE : BLADE :: (A) cut : scissors
(B) strike : bell (C) reduce : fat
(D) focus : image (E) flatten : hill

24. DEBATE:FORENSIC ::

(A) decathlon:athletic
(B) applause:dramatic
(C) controversy:scientific
(D) anthology:descriptive
(E) diagnosis:experimental

25. ANNEX:BUILDING :: (A) tenure:office
(B) rider:document (C) lease:apartment
(D) chapter:text (E) limb:extremity

GO ON TO THE NEXT PAGE

3

Each passage below is followed by questions based on its content. Answer all questions following a passage on the basis of what is stated or implied in that passage.

(The passages for this test have been adapted from published material. The ideas contained in them do not necessarily represent the opinions of the College Board or Educational Testing Service.)

One of the favorite operations of the alchemists was a process they called "calcining." It consisted of placing a mineral or a metal into a crucible and heating it until it was completely converted into a powdery solid. Since this powdery solid evidently resisted prolonged heating better than anything else, it was believed to have special "virtues" and was called "calx," as for example, the calx of arsenic. Calcining, incidentally, provided further proof of the superiority of gold because gold did not have a calx.

In 1489 a German alchemist, Eck von Sulzbach, made the startling discovery that the calx of a metal weighed more than the metal before calcining. The only explanation he could think of was that a "ghost" had come from somewhere and moved into the calx, thereby increasing its weight.

We now know that Eck von Sulzbach had come close to the truth. Calcining a metal means, in present-day terminology, to oxidize it with the oxygen in the atmosphere, and naturally the compound of metal plus oxygen will weigh more than the metal itself. But the fifteenth-century alchemist could not arrive at this conclusion because Aristotle, whose word was still undisputed, had said that air was an element that did not combine with metals. However, the time was not far off when this statement was to be questioned.

26. The passage primarily answers which of the following questions?
    - (A) Why was calcining a favorite process of the alchemists?
    - (B) Who was Eck von Sulzbach and how has he influenced the progress of science?
    - (C) To what extent were alchemists influenced by the teachings of Aristotle?
    - (D) What is calcining and how did the alchemists interpret it?
    - (E) How did other alchemists react to Eck von Sulzbach's discovery of a "ghost"?

27. According to the passage, the calx of two ounces of arsenic will weigh
    - (A) less than one ounce
    - (B) one ounce
    - (C) one and one-half ounces
    - (D) two ounces
    - (E) more than two ounces

28. The last sentence of the passage suggests that, after the fifteenth century, some alchemists began to wonder whether
    - (A) the calx of a metal had any special "virtues"
    - (B) air or some part of it could combine with metals
    - (C) Aristotle's description of gold was accurate
    - (D) calcining increased or decreased the value of a metal
    - (E) the calx of metal really weighed more than the metal did before calcining

GO ON TO THE NEXT PAGE

"That's right. My suitcases were on the gangplank and there was Annie on the quay. That was the first time I thought of her at all. Annie is my wife, but she wasn't then. I looked at her in the pouring rain with tears in her eyes and the porters hitting up against her. I got to thinking of the things I'd said to her one time or another. You know the things you're apt to say to a girl?"

"I do," said the young man.

"I tried to get back for a few last words with her but I had to keep standing aside to let people pass.

" 'Come on up, Manny,' shouted the lads from the deck.

" 'Goodbye, Manny,' Annie called in a little voice."

"You didn't go down!"

"Down I went. The boys called me to come back but the wind blew their words away.

" 'I knew you'd come to your senses,' Annie said. 'Where's your fiddle?'

"By God, if I hadn't left the fiddle on deck! I shouted at Timmy Coyne–that's the fellow next to me in the photo. He played the piano in our band, by the way.

" 'The fiddle!' I shouted against the noise and confusion of the boat pulling away. Timmy right away ducks down then ups and rests the fiddle case on the rails. 'Catch!' he shouts.

"A fellow in the crowd next to me leaps to catch it but you know how slippery them wooden boards can be? Well, down slips the fellow and fiddle and case and even the little bow were smashed to smithereens. You should have heard the crowd laughing."

"What did Annie say?"

" 'It's the hand of God,' she said."

29. In the context of the passage, the phrase "That's right" (line 1) suggests primarily that

(A) Manny and the person he is addressing are in complete agreement
(B) the conversation we are about to read is already in progress
(C) Manny is eager to recount his experience to whomever will listen
(D) one story has been told and another is about to begin
(E) Manny wants his listener to believe what he has told him

30. It can be inferred that all of the following contributed to Manny's decision to rejoin Annie on the dock EXCEPT

(A) Annie's tears
(B) the sight of Annie in the rain
(C) the sight of Annie jostled by porters
(D) the weakness of Annie's voice saying goodbye
(E) the attitude of the men on the deck toward Annie

31. In the context of the passage, the fiddle is representative of all of the following EXCEPT

(A) freedom
(B) youth
(C) religious faith
(D) creative imagination
(E) unexplored possibility

GO ON TO THE NEXT PAGE

Those who think of evolution as merely a process of adaptation to particular environments regard the emergence of life from the water as having been made possible by its adaptation to the dry environment outside. Though this hypothesis is not false in itself, it represents too short a view. In the long view, evolution has been toward an ever-increasing independence of the natural external environment, whatever it may be

Let me give an example. To this day, the development of a fertilized animal cell through the embryonic stage can take place only in a liquid environment. Therefore, the first organisms to emerge on land, represented today by frogs, had to return to the water to lay their eggs. What finally brought about complete independence from the water was the development of the amniotic egg, which packaged the watery environment required by the embryo inside a membrane or shell (i.e., an amnion). So the liquid environment that a human embryo requires is provided inside an amnion inside the mother's womb. A human embryo, unlike that of a fish, is independent of the environment external to its mother, indifferent to changes of temperature–indifferent, even, to whether the environment is wet or dry.

The first human beings to emerge from the earth's atmospheric envelope, in the 1960's, were able to do so because they packaged that atmosphere in their amniotic spacesuits and spaceships. Thus, human beings are now able to visit the moon in spite of a lunar environment that would kill them on contact. What these astronauts represent is not adaptation to the natural environment but independence of it.

32. The author's primary purpose in the passage is to
    - (A) reveal the miracle of human reproduction
    - (B) offer a wider interpretation of evolution
    - (C) explain the larger meaning of the word "amnion"
    - (D) illustrate the difference between a frog embryo and a human embryo
    - (E) compare the space age with the emergence of life from the water

33. From the tone and content of the passage it can be inferred that the author is addressing which of the following audiences?
    - (A) Highly specialized scientists
    - (B) Marine biologists
    - (C) Educated nonprofessionals
    - (D) Aspiring astronauts
    - (E) Environmental conservationists

34. The author uses the development of the amniotic egg to illustrate which of the following?
    - (A) The freeing of the reproductive process from its earlier dependence on the external environment
    - (B) The continual dependence of life on a liquid, external environment, despite the emergence of creatures on land
    - (C) The adaptation of organic life to the external environment
    - (D) The independence of the first organisms to emerge on land
    - (E) The adaptive nature of the evolutionary process

35. We can infer from the passage that the author sees spacesuits and spaceships as examples of
    - (A) atmospheric adaptation
    - (B) artificial progress
    - (C) futuristic apparel
    - (D) a technological variation of evolutionary progress
    - (E) an innovative packaging in modern technology

GO ON TO THE NEXT PAGE

Among economists, John Maynard Keynes ranks with Adam Smith and Karl Marx in the influence that his views have exerted on the general public. He had the vision to see that economics lacked a general theory of demand, and he proceeded with boldness and brilliance to construct one.

His theory produced the startling conclusion that highly developed industrial countries suffer from a chronic deficiency of demand and that this deficiency is bound to grow worse as countries become richer. Hence, Keynes called on government to assume a new responsibility and a new function: to close the growing gap between the power of progressive economies to produce and the size of effective private demand. Keynes suggested two general lines of action—that of controlling the size of the gap through changes in the distribution of income and that of offsetting the gap through greater government spending.

Keynes's theory contributed invaluable tools of analysis to economics and started hundreds of able economists in many lands studying the important problems that the theory opened up. No one in the history of economics has done as much as Keynes to stimulate good work. But Keynes's theory has turned out to be wrong in all its essentials. Although intended to be a "general" theory, applicable to all conditions, it was unduly molded by the depressed Thirties, the period during which Keynes composed it. Since then, economists have determined that advanced economies do not suffer from a chronic deficiency of demand—they suffer from a chronic excess of demand. It would be hard today to find an advanced economy that is not struggling to control demand, and most of them are having only partial success.

It is among the undeveloped economies, precisely where Keynes did not expect to find a chronic shortage of demand, that unemployment is endemic and most severe. Keynes's theory that unemployment is caused by an excessive disposition to save obviously does not explain the high unemployment in countries that are too poor to have any savings at all. The high unemployment in undeveloped countries is best explained by Marx's theory of unemployment—that people lack work because savings are insufficient to provide the growing labor force with the tools of production.

Why has Keynes turned out to have been so completely wrong? He made two basic mistakes. In the first place, he assigned to consumers a relatively passive role in determining the demand for goods. In the second place, he overlooked the fact that the development of investment opportunities is itself an expanding industry, able to supply the community with a rapidly growing number of investment outlets.

36. The author is primarily concerned with which of the following?

(A) Rebutting Keynes's theory of demand
(B) Presenting a theoretical approach to the problem of unemployment
(C) Extending the range of influence of Keynesian theory
(D) Contrasting Keynesian economics with Marxist theories
(E) Establishing the economic responsibilities of modern governments

37. It can be inferred that the author attributes the unsoundness of Keynes's theory chiefly to the fact that it

(A) lacks conclusive proof and is untried
(B) rejects a positive course of action
(C) considers as universal some rules that pertain to a specific economic era
(D) allows too many exceptions to its basic concepts
(E) does not concur with theories of contemporary economists

38. Which of the following does the author particularly emphasize as a Keynesian contribution to the field of economics?

(A) Formulation of a new attitude toward government
(B) Analysis of the theory of unemployment
(C) Development of a theory of income distribution
(D) Stimulation of the thought and work of other economists
(E) Expansion of the theories of Smith and Marx

39. The author notes that one of the causes of unemployment that Keynes overlooked is

(A) excessive private demand
(B) insufficient savings
(C) governmental intervention
(D) lack of initiative
(E) mechanization

40. As used in line 39, the word "endemic" can best be interpreted to mean

(A) contagious and increasing
(B) exclusive and persistent
(C) unlikely to be found
(D) difficult to explain
(E) native to the locality

S T O P

IF YOU FINISH BEFORE TIME IS CALLED, YOU MAY CHECK YOUR WORK ON THIS SECTION ONLY.
DO NOT WORK ON ANY OTHER SECTION IN THE TEST.

5

SECTION V

Time—30 minutes

35 QUESTIONS

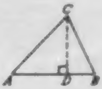
In this section solve each problem, using any available space on the page for scratchwork. Then decide which is the best of the choices given and blacken the corresponding space on the answer sheet.

The following information is for your reference in solving some of the problems.

Circle of radius r: Area $= \pi r^2$; Circumference $= 2\pi r$
The number of degrees of arc in a circle is 360.
The measure in degrees of a straight angle is 180.

Definitions of symbols:

$=$	is equal to	$\leq$	is less than or equal to
$\neq$	is unequal to	$\geq$	is greater than or equal to
$<$	is less than	$\parallel$	is parallel to
$>$	is greater than	$\perp$	is perpendicular to



Triangle: The sum of the measures in degrees of the angles of a triangle is 180.

If $\angle CDA$ is a right angle, then

(1) area of $\triangle ABC = \frac{AB \times CD}{2}$

(2) $AC^2 = AD^2 + DC^2$

Note: Figures which accompany problems in this test are intended to provide information useful in solving the problems. They are drawn as accurately as possible EXCEPT when it is stated in a specific problem that its figure is not drawn to scale. All figures lie in a plane unless otherwise indicated. All numbers used are real numbers.

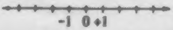
---

1. If $2x + 4 = 9$, then $x - \frac{1}{2} =$

(A) 2 (B) 3 (C) 6 (D) 12 (E) 14

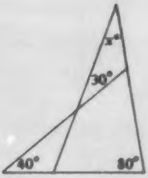
2. $(8\sqrt{4})(2\sqrt{9}) =$

(A) 48 (B) 72 (C) 96 (D) 298 (E) 576



3. On the number line above, a particle starts at 0 and moves a distance of 2 to the right, then from there moves 4 to the left, then 6 to the right, and so on, alternating directions and lengthening each move by 2. At the end of the fifth move, where on the number line will the particle be located?

(A) -8 (B) -4 (C) +1 (D) +6 (E) +10



4. In the figure above, $x =$

(A) 120 (B) 90 (C) 60 (D) 45 (E) 30

5. If x and y are positive integers and if $\frac{x}{y} = 1$ and $(x + y)^2 = z$, which of the following CANNOT equal z?

(A) 4
(B) 9
(C) 16
(D) 36
(E) 64

6. If a \$27,000 prize was divided among three people in the ratio of 2 : 3 : 5, what was the value of the largest share?

(A) \$18,900
(B) \$13,500
(C) \$8,100
(D) \$5,400
(E) \$2,700

7. If y is some number between 4 and 10, which of the following could be the average of the numbers 2, 5, 6, 8, 9, and y?

(A) 4.3
(B) 6.2
(C) 7.8
(D) 9.1
(E) 10.0

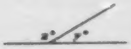
GO ON TO THE NEXT PAGE

Questions 8-27 each consist of two quantities, one in Column A and one in Column B. You are to compare the two quantities and on the answer sheet blacken space

- A if the quantity in Column A is greater;
- B if the quantity in Column B is greater;
- C if the two quantities are equal;
- D if the relationship cannot be determined from the information given.

Notes:
1. In certain questions, information concerning one or both of the quantities to be compared is centered above the two columns.
2. In a given question, a symbol that appears in both columns represents the same thing in Column A as it does in Column B.
3. Letters such as x, n, and k stand for real numbers.

EXAMPLES

	Column A	Column B	Answers
E1.	2×6	$2 + 6$	● Ⓑ Ⓒ Ⓓ
			
E2.	$180 - x$	y	Ⓐ Ⓑ ● Ⓓ
E3.	$p - q$	$q - p$	Ⓐ Ⓑ Ⓒ ●

Column A — Column B

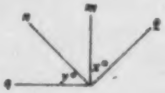
8. $6 + 4^2$ — 20

$x + 5 = 19$
$y + 5 = 14$

9. x — y

$a = b = c = d = e$

10. $a + 2b$ — $c + d + e$



$n \perp \ell$ and $q \perp m$

11. x — y

Column A — Column B

12. -0.013 — -0.087

$x \neq 0$

13. $3x^2$ — $(3x)^2$

$p > q > r > 0$

14. $p^2 + q^2$ — $pq + qr$

TELEPHONE RATES

From P to	First 3 Minutes	Each Additional Minute
Q	45¢	10¢
S	15¢	5¢

15. The cost of a 20-minute call from P to S — $1.00

$xyz > 0$
$x < 0$

16. y — z

16% of $2x = 96$.

17. 16% of x — 48

GO ON TO THE NEXT PAGE

SUMMARY DIRECTIONS FOR COMPARISON QUESTIONS

Answer: A if the quantity in Column A is greater;
B if the quantity in Column B is greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

$-7 < n < -3$

	Column A	Column B
18.	8	n^2
19.	Volume of a cylinder with radius 2	Volume of a cylinder with radius 4

$x + y = 8$
$x - y = 12$

20.	y	0

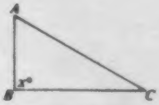
A car travels d kilometers in t minutes at a constant rate of 60 kilometers per hour.

21.	t	d

LIST I: 1, 2, 3
LIST II: 2, 3, 4

x is a number chosen at random from List I and y is a number chosen at random from List II.

22.	The most likely value of $x + y$	5



Note: Figure not drawn to scale.

$x \neq 90$

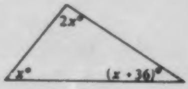
23.	$AB^2 + BC^2$	AC^2

x is an integer greater than 1. $\boxed{x}$ denotes the smallest positive integer factor of x not equal to 1.

	Column A	Column B
24.	$\boxed{x}$	$\boxed{x^2}$

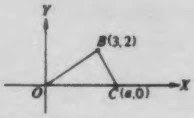
n is a positive integer.

25.	$(-1)^n$	$[1 + (-1)]^n$



Note: Figure not drawn to scale.

26.	$2x$	$x + 36$



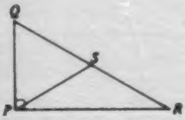
Note: Figure not drawn to scale.

ΔOBC has an area of 12.

27.	a	8

GO ON TO THE NEXT PAGE

Solve each of the remaining problems in this section using any available space for scratchwork. Then decide which is the best of the choices given and blacken the corresponding space on the answer sheet.



28. In the figure above, if $\triangle PQS$ is equilateral, what is the ratio $\frac{PS}{SR}$?

(A) $\frac{1}{2}$ (B) $\frac{1}{\sqrt{3}}$ (C) $\frac{\sqrt{3}}{2}$ (D) 1 (E) 2

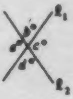
29. The difference between $6\frac{2}{3}$ hours and $7\frac{3}{5}$ hours is how many minutes?

(A) 40 (B) 45 (C) 50 (D) 56 (E) 64

30. For $a \neq 0$, $\frac{a^{x+y}}{a^x} =$

(A) a^y (B) $\frac{1}{a^y}$ (C) $-a^y$

(D) a^{1+y} (E) $1 + a^y$



31. In the figure above, which of the following must be equal to 1 ?

I. $\frac{a+b}{c+d}$

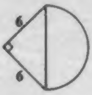
II. $\frac{ab}{cd}$

III. $\frac{c+d-a}{b}$

(A) None (B) I only (C) II only
(D) I and II only (E) I, II, and III

32. The possible scores on a weekly math quiz range from 0 to 100. If Pat receives scores of 65 and 75 on the first two quizzes, what is the lowest possible score Pat can receive on the third in order to average 75 on the first 4 quizzes?

(A) 50
(B) 55
(C) 60
(D) 65
(E) 70

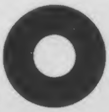


33. In the figure above, the area of the semicircle is

(A) 36π (B) 18π (C) 12π (D) 9π (E) $\frac{9}{2}\pi$

34. There are fewer than 40 students enrolled in a certain class. If, at a certain time, $\frac{2}{9}$ of the students are absent from school and $\frac{1}{4}$ of those in school have gone to the library, what is the total class enrollment?

(A) 18
(B) 27
(C) 28
(D) 32
(E) 36



35. In the figure above, what is the greatest number of nonoverlapping regions into which the shaded region can be divided with exactly two straight lines?

(A) 6 (B) 5 (C) 4 (D) 3 (E) 2

S T O P

IF YOU FINISH BEFORE TIME IS CALLED, YOU MAY CHECK YOUR WORK ON THIS SECTION ONLY.
DO NOT WORK ON ANY OTHER SECTION IN THE TEST.

SECTION VI

Time—30 minutes

50 QUESTIONS

This section tests your ability to use standard written English, the kind of English found in most college textbooks.

Directions: The following sentences contain problems in grammar, usage, diction (choice of words), and idiom.

Some sentences are correct.
No sentence contains more than one error.

You will find that the error, if there is one, is underlined and lettered. Assume that all other elements of the sentence are correct and cannot be changed. In choosing answers, follow the requirements of standard written English.

If there is an error, select the one underlined part that must be changed to make the sentence correct, and blacken the corresponding space on your answer sheet.

If there is no error, blacken answer space Ⓔ.

EXAMPLE:

He spoke bluntly and angrily to we
B C
spectators. No error
D E

SAMPLE ANSWER

Ⓐ Ⓑ ● Ⓓ Ⓔ

1. Most people listen to the weather forecast every day,
but they know hardly nothing about the forces
A B C
that influence the weather. No error
D E

2. Him and the other delegates immediately accepted
A B C
the resolution drafted by the neutral states.
D
No error
E

3. The foundations of psychoanalysis were established
A
by Sigmund Freud, who begun to develop his
B C
theories in the 1880's. No error
D E

4. Charter flights, although much less expensive than
A B
scheduled travel, is well known for late takeoffs
C
and other inconveniences. No error
D E

5. Many women reenter the job market when their
A B
youngest children started kindergarten. No error
C D E

6. After facing so many groups of angry people,
A
Martin Luther King, Jr. became accustomed to
B
confrontations that were unpredictable and even
C D
potentially explosive. No error
E

GO ON TO THE NEXT PAGE

7. $\underset{A}{\underline{\text{During}}}$ the early Middle Ages, before the development of the printing press, $\underset{B}{\underline{\text{virtually}}}$ the only books were $\underset{C}{\underline{\text{those}}}$ that $\underset{D}{\underline{\text{are laboriously}}}$ copied by monks. $\underset{E}{\underline{\text{No error}}}$

8. The decision that $\underset{A}{\underline{\text{has just been}}}$ $\underset{B}{\underline{\text{agreed with}}}$ by the committee members should serve as a basis $\underset{C}{\underline{\text{for their}}}$ work in the $\underset{D}{\underline{\text{years to come}}}$. $\underset{E}{\underline{\text{No error}}}$

9. In her novels, Nella Larson $\underset{A}{\underline{\text{focused on}}}$ the problems of young black women $\underset{B}{\underline{\text{which}}}$ $\underset{C}{\underline{\text{lived in}}}$ Europe and America $\underset{D}{\underline{\text{during}}}$ the 1920's. $\underset{E}{\underline{\text{No error}}}$

10. As general supervisor, Ms. Rodríguez $\underset{A}{\underline{\text{expects}}}$ her staff $\underset{B}{\underline{\text{to work}}}$ as carefully $\underset{C}{\underline{\text{and as hard}}}$ $\underset{D}{\underline{\text{the way}}}$ she herself does. $\underset{E}{\underline{\text{No error}}}$

11. People who dislike cats $\underset{A}{\underline{\text{sometimes}}}$ criticize them $\underset{B}{\underline{\text{for being}}}$ aloof and independent; people who are $\underset{C}{\underline{\text{fond of}}}$ cats often admire $\underset{D}{\underline{\text{them for}}}$ the same qualities. $\underset{E}{\underline{\text{No error}}}$

12. $\underset{A}{\underline{\text{Only a few feet}}}$ $\underset{B}{\underline{\text{beyond}}}$ those abandoned $\underset{C}{\underline{\text{and crumbling}}}$ buildings $\underset{D}{\underline{\text{stand}}}$ a beautiful cluster of weeping willow trees. $\underset{E}{\underline{\text{No error}}}$

13. $\underset{A}{\underline{\text{In contrasting to}}}$ the popular view, the characteristic way of life $\underset{B}{\underline{\text{of many}}}$ Native American peoples $\underset{C}{\underline{\text{was}}}$ neither nomadic $\underset{D}{\underline{\text{nor}}}$ warlike. $\underset{E}{\underline{\text{No error}}}$

14. $\underset{A}{\underline{\text{Until recently}}}$, Americans $\underset{B}{\underline{\text{drank}}}$ five cups of coffee $\underset{C}{\underline{\text{for every}}}$ cup of tea, but now they are drinking more $\underset{D}{\underline{\text{of it}}}$. $\underset{E}{\underline{\text{No error}}}$

15. All our neighbors are wondering $\underset{A}{\underline{\text{what}}}$ the government $\underset{B}{\underline{\text{will do}}}$ about the $\underset{C}{\underline{\text{raising}}}$ gasoline prices and the $\underset{D}{\underline{\text{shortage}}}$ of oil. $\underset{E}{\underline{\text{No error}}}$

16. There $\underset{A}{\underline{\text{are}}}$ certain areas $\underset{B}{\underline{\text{of}}}$ the southern United States $\underset{C}{\underline{\text{where}}}$ snow $\underset{D}{\underline{\text{has never fallen}}}$. $\underset{E}{\underline{\text{No error}}}$

17. $\underset{A}{\underline{\text{When}}}$ Pelé $\underset{B}{\underline{\text{was playing}}}$ soccer, he was as talented as, or $\underset{C}{\underline{\text{more talented than}}}$, $\underset{D}{\underline{\text{baseball or basketball}}}$. $\underset{E}{\underline{\text{No error}}}$

18. The needs of children $\underset{A}{\underline{\text{often}}}$ are not $\underset{B}{\underline{\text{taken}}}$ $\underset{C}{\underline{\text{into account}}}$ in the design of public buildings $\underset{D}{\underline{\text{nor}}}$ residential areas. $\underset{E}{\underline{\text{No error}}}$

GO ON TO THE NEXT PAGE

6

19. In a recent created wildlife refuge, the birds
A
are now building their nests, thereby
B C
greatly changing the ecology of the area.
D
No error
E

20. Because flashes of light distract them, entertainers
A
often request that people with a camera refrain from
B C D
taking pictures during performances. No error
E

21. Color blindness may handicap one in some
A
respects, but you can still function normally
B C
in almost all situations you encounter. No error
D E

22. Rising at a rate of ten inches per hour, the Red River
A B
soon posed a serious threat to the residents of the
C D
valley. No error
E

23. Astronomy and astrology are both concerned in
A B
the heavenly bodies, but their purposes and
C
methods are quite different. No error
D E

24. As the discussion between the two candidates
A
continued, they spoke more and more loudly
B C
and with less and less dignity. No error
D E

25. The doctor lived in the tropics for three years
A B
before he was inflicted by malaria. No error
C D E

GO ON TO THE NEXT PAGE

The following sentences test correctness and effectiveness of expression.

Directions: In each of the following sentences, some part or all of the sentence is underlined. Under each sentence you will find five ways of phrasing the underlined part. Select the answer that produces the most effective sentence, one that is clear and exact, without awkwardness or ambiguity, and blacken the corresponding space on your answer sheet. In choosing answers, follow the requirements of standard written English, but do not make a choice that changes the meaning of the original sentence.

Answer (A) is always the same as the underlined part. Choose answer (A) if you think the original sentence needs no revision.

EXAMPLE:

Ms. Rose planning to teach a course in biology next summer.

(A) planning
(B) are planning
(C) have planned
(D) with a plan
(E) plans

SAMPLE ANSWER

Ⓐ Ⓑ Ⓒ Ⓓ ●

26. Because dodo birds could not fly, so they were killed by the hogs and monkeys brought to the islands by the explorers.

(A) fly, so they were killed
(B) fly, they were killed
(C) fly and they were killed
(D) fly, and this allowed them to be killed
(E) fly, killing them

27. Performing before an audience for the first time, fear suddenly overcame the child and she could not remember her lines.

(A) fear suddenly overcame the child and she could not remember her lines
(B) the lines could not be remembered by the child because she was overcome by fear
(C) the child was suddenly overcome by fear and could not remember her lines
(D) the child was suddenly overcome by fear, she could not remember her lines
(E) suddenly the child was overcome by fear, and consequently not remembering her lines

28. Young people are not rejecting marriage, but some postponing it.

(A) some postponing it
(B) some are postponing it
(C) it is postponed by some of them
(D) it is being postponed
(E) some having postponed it

29. Violin makers know that the better the wood is seasoned, the better the results for the tone of the instrument.

(A) better the results for the tone of the instrument
(B) better the tone of the instrument
(C) better the result is for the instrument's tone
(D) resulting tone will be better
(E) result will be a better instrument tone

30. Although today many fabrics are made from synthetic fibers, at one time all natural fibers were used in their manufacture.

(A) all natural fibers were used in their manufacture
(B) all fabrics were made of natural fibers
(C) they were making them all of natural fibers
(D) they made fabrics all of natural fibers
(E) their manufacture was of all natural fibers

31. Between three and four per cent of all children born with hearing defects serious enough to require medical treatment.

(A) born with hearing defects
(B) being born with hearing defects that are
(C) are born with hearing defects
(D) are born with hearing defects, these are
(E) born with hearing defects which are

GO ON TO THE NEXT PAGE

6

32. Issued in Great Britain in 1840, the first gummed postage stamp in history was known as the "Penny Black."
   (A) the first gummed postage stamp in history was known as
   (B) they called the first gummed postage stamp in history
   (C) history refers to the first gummed postage stamp as
   (D) was the first gummed postage stamp in history,
   (E) the first gummed postage stamp in history being known as

33. A number of parents are concerned about the protection of high school athletes and the many injuries being reported.
   (A) athletes and the many injuries being reported
   (B) athletes and numerous reports of injuries
   (C) athletes because of reports of numerous injuries
   (D) athletes, but many injuries are being reported
   (E) athletes, numerous injuries have been reported

34. Many inferior films earn a great deal of money for their producers, some extremely good ones do not.
   (A) Many inferior films earn
   (B) Many an inferior film earns
   (C) With many inferior films which earn
   (D) However, many inferior films earn
   (E) Although many inferior films earn

35. Gas rationing would force consumers to use their cars less, use public transportation more, while conserving gasoline.
   (A) while conserving
   (B) as well as conserving
   (C) conserving
   (D) and thereby conserve
   (E) to conserve

36. Dolphins have a basic social organization, a system of communication, and their brains are highly developed.
   (A) and their brains are highly developed
   (B) and highly developed brains
   (C) with highly developed brains
   (D) while their brains are developed highly
   (E) but their brains are developed highly

37. The Romans built many magnificent aqueducts and roads, they were truly the great engineers of the ancient world.
   (A) roads, they were truly the great
   (B) roads, consequently they were truly the great
   (C) roads and as a consequence, they were truly the greatest
   (D) roads; truly the great
   (E) roads; they were truly the great

38. The Japanese began their remarkable postwar recovery, but they were able to recover from the psychological effects of the war.
   (A) recovery, but they were
   (B) recovery, where they were
   (C) recovery only when they were
   (D) recovery only when being
   (E) recovery because it was

39. For success in sports, it is important not only to be in good physical condition but also to know the rules of the game.
   (A) to be in good physical condition but also to know
   (B) being in good physical condition but also to know
   (C) to be in good physical condition but also knowing
   (D) being in good physical condition but also knowing
   (E) that one be in good physical condition but also that you know

40. Founded in 1910, the Urban League was established and has continuously had for its purpose the aim of assisting blacks who live in large cities.
   (A) was established and has continuously had for its purpose the aim of
   (B) was established to aim in the purpose of
   (C) has continuously had for its purpose the aim of
   (D) has always aimed at
   (E) was with the purpose of

GO ON TO THE NEXT PAGE

351

Note: The remaining questions are like those at the beginning of the section.

Directions: In each sentence, select the one underlined part that must be changed to make the sentence correct, and blacken the corresponding space on your answer sheet.

If there is no error, blacken answer space Ⓔ.

EXAMPLE:

He spoke $\underset{A}{\underline{\text{bluntly}}}$ and $\underset{B}{\underline{\text{angrily}}}$ to $\underset{C}{\underline{\text{we}}}$ $\underset{D}{\underline{\text{spectators.}}}$ $\underset{E}{\underline{\text{No error}}}$

SAMPLE ANSWER

Ⓐ Ⓑ ● Ⓓ Ⓔ

41. Idealists are $\underset{A}{\underline{\text{not always}}}$ $\underset{B}{\underline{\text{as ignorant of}}}$ realities as $\underset{C}{\underline{\text{his}}}$ critics would $\underset{D}{\underline{\text{like to}}}$ believe. $\underset{E}{\underline{\text{No error}}}$

42. $\underset{A}{\underline{\text{Once an}}}$ Italian colony, Eritrea $\underset{B}{\underline{\text{was captured}}}$ by the British in 1941 and was $\underset{C}{\underline{\text{united with}}}$ Ethiopia in $\underset{D}{\underline{\text{the early}}}$ 1950's. $\underset{E}{\underline{\text{No error}}}$

43. $\underset{A}{\underline{\text{More women}}}$ are competing $\underset{B}{\underline{\text{in athletics}}}$ $\underset{C}{\underline{\text{as}}}$ new opportunities for training develop and as sports programs $\underset{D}{\underline{\text{will expend.}}}$ $\underset{E}{\underline{\text{No error}}}$

44. Mayor Jameson $\underset{A}{\underline{\text{was eager}}}$ $\underset{B}{\underline{\text{to demonstrate}}}$ that she was not prejudiced in favor of any particular political groups $\underset{C}{\underline{\text{and}}}$ residents $\underset{D}{\underline{\text{of}}}$ specific areas of the city. $\underset{E}{\underline{\text{No error}}}$

45. $\underset{A}{\underline{\text{In any lively}}}$ debate, there $\underset{B}{\underline{\text{is}}}$ at least two $\underset{C}{\underline{\text{sides to}}}$ an issue and five arguments $\underset{D}{\underline{\text{for every}}}$ point of view presented. $\underset{E}{\underline{\text{No error}}}$

46. After she $\underset{A}{\underline{\text{had spoke}}}$ for fifteen minutes, the senator answered $\underset{B}{\underline{\text{candidly}}}$ the questions asked by the reporters, $\underset{C}{\underline{\text{whose}}}$ skepticism $\underset{D}{\underline{\text{was}}}$ obvious. $\underset{E}{\underline{\text{No error}}}$

47. Demonstrating new products $\underset{A}{\underline{\text{is}}}$ a more expensive $\underset{B}{\underline{\text{but more effective}}}$ sales technique than $\underset{C}{\underline{\text{merely talking}}}$ about $\underset{D}{\underline{\text{them.}}}$ $\underset{E}{\underline{\text{No error}}}$

48. Moscow's city officials plan $\underset{A}{\underline{\text{doubling}}}$ the size of the city's subway system, which $\underset{B}{\underline{\text{already includes}}}$ $\underset{C}{\underline{\text{more than}}}$ one hundred $\underset{D}{\underline{\text{miles of}}}$ track. $\underset{E}{\underline{\text{No error}}}$

49. The speaker argued that neither industry $\underset{A}{\underline{\text{nor}}}$ government $\underset{B}{\underline{\text{have taken}}}$ full responsibility $\underset{C}{\underline{\text{for reducing}}}$ the spiral $\underset{D}{\underline{\text{of inflation.}}}$ $\underset{E}{\underline{\text{No error}}}$

50. The average taxpayer waits until April $\underset{A}{\underline{\text{to file}}}$ a tax return $\underset{B}{\underline{\text{unless}}}$ $\underset{C}{\underline{\text{they expect}}}$ to receive a refund $\underset{D}{\underline{\text{of considerable size.}}}$ $\underset{E}{\underline{\text{No error}}}$

IF YOU FINISH BEFORE TIME IS CALLED, YOU MAY CHECK YOUR WORK ON THIS SECTION ONLY.
DO NOT WORK ON ANY OTHER SECTION IN THE TEST.

HOW TO SCORE YOUR SAMPLE TEST

SAT scores are reported on the College Board scale of 200 to 800. A score of 200 isn't necessarily equal to zero, nor does a score of 800 mean that you answered every question correctly. The TSWE score is reported on a scale of 20 to 60+ . The TSWE is not intended to distinguish among students whose command of English is better than average. Therefore, students who receive a score of 60+ can conclude that the test is not difficult enough to measure the upper limits of their ability.

There are two steps in scoring:

1. determining verbal, mathematical, and TSWE raw scores (counting the number of right responses and subtracting a fraction of the number wrong), and
2. finding the College Board scaled scores for the corresponding raw scores.

You will understand the scoring process better if you try to apply it to the sample test you just took.

Determining Your Raw Score

For the SAT-verbal sections and the TSWE

The 85 questions in sections 1 and 3 yield the verbal score. Check your answers against the table on page 45. Scan the verbal questions (sections 1 and 3) and count the number of right answers and the number of wrong answers. Count one point for each correct response. For each wrong response, subtract one-fourth of a point. There is no need to count the questions you omitted because they do not affect your scores.

Raw score = number right minus one-fourth of the number wrong

For example, if you have 46 right responses to the verbal questions in sections 1 and 3 and 31 wrong responses, then your verbal raw score is 46 minus one-fourth of 31, which equals 38. Or $46 - \frac{1}{4}(31) = 38\frac{1}{4}$, which rounds down to 38 (the nearest integer).

You can find your TSWE score in the same way. For example, if you have 39 right responses to the TSWE questions in section 6 and 10 wrong responses, then your TSWE raw score is 39 minus one-fourth of 10, which equals 37. Or $39 - \frac{1}{4}(10) = 36\frac{1}{2}$, which rounds up to 37.

For the SAT-mathematical sections

Determining the mathematical score is a little more complicated because some of the questions contain only four possible answer choices instead of the usual five. Count the 20 four-choice questions (numbers 8 through 27 in section 5) as a separate score. Count the 40 five-choice mathematical questions (all of section 2 and numbers 1 through 7 and 28 through 35 in section 5) as a second mathematical score. Obtain the raw scores as follows:

Raw score for 20 four-choice questions = number right minus one-third of the number wrong

Raw score for 40 five-choice questions = number right minus one-fourth of the number wrong

Add the results of the two scores to obtain one total mathematical raw score. For example, suppose you have 10 right and 10 wrong for the 20 four-choice questions; then, the raw score is 10 minus one-third of 10, which equals $6\frac{2}{3}$. Or $10 - \frac{1}{3}(10) = 6\frac{2}{3}$. And suppose you have 24 right and 11 wrong for the 40 five-choice questions; then the raw score is 24 minus one-fourth of 11, which equals $21\frac{1}{4}$. Or $24 - \frac{1}{4}(11) = 21\frac{1}{4}$.

To obtain the total raw score, add $6\frac{2}{3}$ to $21\frac{1}{4}$ for a mathematical score of 28. (Round any scores with fractions to whole numbers only after the two parts are summed.)

Finding Your College Board Score Range

Use the table on page 46 to find the range of College Board scores for the SAT-verbal, SAT-mathematical, and TSWE raw scores you obtained on the sample test. For example, in the column headed "Raw Score," find your verbal raw score or the score nearest to it. In the column headed "SAT-verbal," find the range of College Board scores for it. Let's say your verbal raw score is 38. The number closest to it is 40. The column headed "SAT-verbal" shows that the range of College Board scores for a raw score of 40 is 460-480. The range for your score of 38 would be somewhat lower.

The table shows a range of scores because there are slight variations in the difficulty of different versions of the test. On an easier version of the test, the raw score obtained would result in the College Board score that is at the lower end of the range. For a more difficult version, the raw score would result in the College Board score that is at the higher end of the range.

Let's take another example. Suppose your mathematical raw score is 27. The number closest to it in the column headed "Raw Score" is 25. The column headed "SAT-mathematical" shows that the range of College Board scores for a raw score of 25 is 470-490. The range for your score of 27 would be somewhat higher.

ANSWERS TO SAMPLE TEST QUESTIONS AND PERCENTAGE OF STUDENTS ANSWERING EACH QUESTION CORRECTLY

Section 1—Verbal			Section 2—Mathematical			Section 3—Verbal			Section 5—Mathematical			Section 6—TSWE		
Question number	Correct answer	Percentage of students answering the question correctly	Question number	Correct answer	Percentage of students answering the question correctly	Question number	Correct answer	Percentage of students answering the question correctly	Question number	Correct answer	Percentage of students answering the question correctly	Question number	Correct answer	Percentage of students answering the question correctly
1	C	96%	1	B	80%	1	E	96%	1	A	78%	1	B	90%
2	B	90	2	B	92	2	A	85	2	C	79	2	A	94
3	D	80	3	E	84	3	B	87	3*	D		3	B	90
4	E	80	4	C	60	4	C	69	4	E	71	4	C	78
5	A	71	5	C	76	5	D	35	5	B	67	5	D	86
6	B	75	6	D	76	6	A	22	6	B	65	6	E	88
7	C	66	7	E	66	7	B	19	7	B	59	7	D	74
8	B	68	8	C	80	8	A	16	8	A	92	8	B	64
9	D	55	9	B	63	9	A	20	9	A	95	9	B	80
10	E	32	10	B	44	10	E	15	10	C	80	10	D	74
11	A	24	11	B	55	11	C	84	11	C	70	11	E	88
12	A	16	12	C	52	12	E	66	12	A	80	12	D	64
13	C	13	13	A	60	13	D	54	13	B	78	13	A	52
14	A	18	14	C	56	14	D	26	14	A	65	14	D	72
15	B	10	15	C	39	15	A	16	15	C	62	15	C	70
16	C	92	16	A	49	16	B	91	16	D	63	16	E	76
17	A	73	17	E	33	17	D	81	17	C	48	17	D	86
18	D	46	18	D	38	18	D	76	18	B	65	18	D	65
19	B	51	19	B	16	19	B	64	19	D	38	19	A	52
20	A	42	20	A	21	20	C	71	20	B	44	20	C	43
21	E	50	21	D	20	21	D	71	21	C	40	21	A	50
22	A	41	22	B	34	22	E	26	22	C	29	22	E	63
23	D	57	23	A	9	23	D	23	23	D	31	23	B	57
24	B	67	24	D	10	24	A	24	24	C	8	24	E	47
25	C	35	25	D	11	25	B	11	25	D	19	25	D	44
26	C	53				26	D	79	26	C	24	26	B	92
27	E	53				27	E	85	27	A	14	27	C	66
28	D	55				28	B	81	28	D	51	28	B	88
29	C	41				29	B	67	29	D	52	29	B	50
30	A	56				30	E	73	30	A	39	30	B	60
31	D	78				31	C	35	31	E	30	31	C	74
32	C	50				32	B	66	32	C	24	32	A	67
33	A	31				33	C	72	33	D	11	33	C	59
34	E	27				34	A	28	34	E	10	34	E	70
35	C	13				35	D	53	35	B	7	35	D	56
36	E	81				36	A	49				36	B	86
37	B	71				37	C	41				37	E	53
38	B	62				38	D	34				38	C	60
39	D	64				39	B	47				39	A	62
40	A	47				40	E	8				40	D	29
41	E	23										41	C	67
42	C	24										42	E	72
43	B	18										43	D	75
44	D	24										44*	C	
45	B	15										45	B	48
												46	A	88
												47	E	61
												48	A	55
												49	B	42
												50	C	44

Notes: The percentages are based on the analysis of the answer sheets for a random sample of students who took this test in April 1981 and whose mean scores were 430 on the SAT-verbal sections, 470 on the SAT-mathematical sections, and 43 on the TSWE.

The order of the six SAT sections varies from one edition to another. [illegible] Section 4, the experimental section in the edition reprinted in this booklet, is omitted (see page 19).

*Question 3 in Section 5 and question 44 in Section 6 in the edition of the SAT administered on April 4, 1981, were not counted in computing scores. In the sample test in this booklet, these two questions are printed as they appear in a later edition of the SAT.

When you take the SAT, your score is likely to differ from the score you obtained on the sample test. People perform at different levels at different times for reasons unrelated to the test itself. In addition, the precision of any test is limited because it represents only a sample of all the possible questions that could be asked.

RAW SCORES CONVERTED TO SCALED SCORES

Raw Score*	College Board Score Ranges SAT-verbal (85 questions)	SAT-mathematical (60 questions)	TSWE (50 questions)
85	800		
80	750-770		
75	710-730		
70	670-690		
65	630-650		
60	590-620	800	
55	560-580	740-760	
50	530-550	690-720	60+
45	490-510	640-670	57-60+
40	460-480	600-630	52-56
35	420-440	560-590	47-50
30	390-410	510-530	42-46
25	350-370	470-490	37-39
20	320-340	430-450	31-34
15	280-310	380-400	26-29
10	250-270	340-360	21-24
5	210-240	290-320	20
0	200-210	250-280	20

*Raw score = number right minus a fraction of the number wrong

Note: The ranges for the College Board scores are based on all editions of the SAT given from November 1977 through June 1980. Scores for the sample test in this booklet fall within these ranges.

In addition to correct answers to the sample test, the table on page 45 gives percentages, which are based on the analysis of the answer sheets for a random sample of students who took this test in April 1981 and whose mean scores were 430 on the SAT-verbal sections, 470 on the SAT-mathematical sections, and 43 on the TSWE. The percentages will give you an idea of how hard each question was for students with about average verbal or mathematical ability.

For example, the percentage for question 6 in verbal section 1 indicates that 75 percent of a group of students scoring 430 would probably answer this item correctly. The question, therefore, is not very hard for them. Question 10 would be much harder for such a group. Only 32 percent got question 10 right.

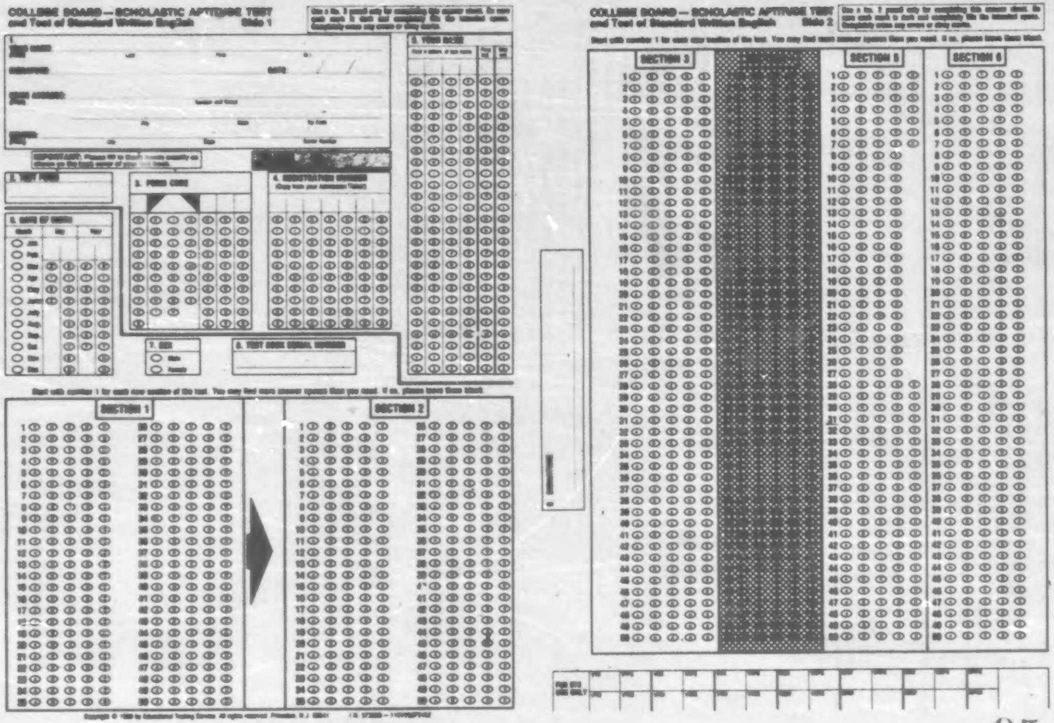
After you have scored your test, analyze your performance.

- For instance, did you omit questions because you ran out of time before you reached the end of a section? Reread pages 3 through 18. The suggestions in them may help you pace yourself better.
- Did you take so long reading directions that you didn't have enough time to answer questions? If you become thoroughly familiar with the test directions reprinted in this booklet, you won't have to spend much time reading them during the test.
- Look at the specific questions you missed. Did you get caught by a choice that was only partly correct? Figure out what step you overlooked in your reasoning.

RECEIVING YOUR SCORE REPORT

About six weeks after you take the SAT and TSWE, you will receive a report that includes your scores and percentile ranks. The percentile ranks show how your scores compare with the scores of certain other groups of students who have taken the SAT and TSWE. With the report, you'll receive a booklet, *Your Student Report*, which will help you interpret and understand the scores, percentile ranks, and other information that appear on your report.The booklet will also explain how colleges use your scores. Your counselor will have additional interpretive material.

We hope that *Taking the SAT* has provided information that will increase your confidence when you take the test. Your suggestions and comments concerning the contents of the booklet will be welcome. Send them to College Board ATP, Box 592, Princeton, NJ 08541.



COLLEGE BOARD — SCHOLASTIC APTITUDE TEST
and Test of Standard Written English Side 1

SECTION 1

SECTION 2

COLLEGE BOARD — SCHOLASTIC APTITUDE TEST
and Test of Standard Written English Side 2

SECTION 3

SECTION 5

SECTION 6

Note: This is a reduced sample of both sides of an actual SAT answer sheet. Do not use Section 4 of this sample answer sheet because that section is omitted from the sample test.

A Message to Students

The question is frequently asked: What can I do about raising my SAT scores or about making them better than they would be otherwise? The answer is: Quickly and immediately probably not much; over longer periods it depends upon how much time, effort, and concentration goes into the preparation.

The Scholastic Aptitude Test measures the extent to which your reasoning ability and skills with words and mathematical concepts have been developed up to the time you take the test. These are abilities that are related to academic success in college and that grow over a lifetime through learning experiences such as those in the family, in school, with your friends and associates, and in reading and independent study. The best preparation for the SAT is to have had varied opportunities of this kind and to have made the most of them.

The skills and abilities the SAT tests tend to grow relatively slowly and at different rates for different people. Whether you have more or less of these abilities does not say anything about your worth as an individual. Many other individual qualities not measured by the SAT, such as motivation, creativity, and artistic skills, have much to do with your sense of satisfaction and your success in life.

If you or your parents have been thinking about special preparation for the SAT outside your regular classroom activities, these six points are worth remembering:

1. The SAT measures developed verbal and mathematical reasoning abilities that are involved in successful academic work in college; it is not a test of some inborn and unchanging capacity.

2. Scores on the SAT are subject to improvement as educational experience, both in and out of school, causes these verbal and mathematical abilities to develop.

3. Development of these abilities is related to the time and effort spent; short-term drill and cramming are likely to have little effect; longer-term preparation that develops skills and abilities can have greater effect.

4. While drill and practice on sample test questions generally result in little effect on test scores, preparation of this kind can familiarize you with different question types and may help to reduce anxiety about what to expect. You can help yourself to become familiar with the test by using the explanations and full sample test in this booklet.

5. Whether longer preparation, apart from that available to you within your regular high school courses, is worth the time, effort, and money is a decision you and your parents must make for yourselves; results seem to vary considerably from program to program, and for each person within any one program. Studies of special preparation programs carried on in many high schools show various results averaging about 10 points for the verbal section and 15 points for the mathematical over and above the average increases that would otherwise be expected from intellectual growth and practice. In other programs results have ranged from virtually no improvement in scores to average gains as high as 25-30 points for particular groups of students or particular programs. Recent studies of commercial coaching have shown a similar range of results. You should satisfy yourself that the results of a special program or course are likely to make a difference in relation to your college admissions plans.

6. Generally, the soundest preparation for the SAT is to study widely with emphasis on academic courses and extensive outside reading. SAT score increases of 20-30 points correspond to about three additional questions answered correctly. Such a result might be obtained by independent study in addition to regular academic course work.

1101608 . DO31P3668 . 200617 . Printed in U.S.A.

Attachment B

The
American
School Board
Journal.

Standardized testing has become education's latest scapegoat

Reprinted with permission, from The American School Board Journal, July 1981. Copyright 1981, the National School Boards Association. All Rights Reserved.

As E.T.S. sees it

Standardized testing has become education's latest scapegoat

By Scarvia B. Anderson

WE AMERICANS have a long history of explaining away any dissatisfaction with our education system by blaming successively vulnerable parties. First we blamed the students, then we blamed their parents (indicting Dr. Spock along the way), then we blamed the teachers, and of course administrators always are blamed for everything. But the late 1970s brought more bad news. Student scores on the Scholastic Aptitude Test (S.A.T.) continued to decline; many volunteers in the armed forces were labeled incompetent in basic skills; colleges began to offer more remedial courses; the crime rate went up—all of this in spite of unprecedented societal attention paid to education legislation and innovation.

Clearly, the time was right to nominate a new villain for the "what's-wrong-with-our-schools?" drama. No group was more eager to serve as the casting agent than those who most recently had been featured players—teachers, or more specifically, one of the two major teacher unions. Without so much as an audition, the National Education Association (N.E.A.) quickly picked *standardized testing* as the villain. The choice was praised by the National Association for the Advancement of Colored People, Ralph Nader, several state legislators, and two social scientists at the Harvard Medical School.

Their arguments against testing were patterned after the logic of Lewis Carroll, author of *Alice's Adventures in Wonderland*: If minorities scored lower on tests, *if* test scores were going down, if students' scores on tests could be improved, then the trouble lay with the tests, not with the schools. "If only the tests were cleared away," they said, "it would be grand."

But tests, of course, have not been "cleared away"—*nor* are they likely to be. Educators would be wise to recognize this fact and to redirect their energies from damning tests to learning as much as they can about them (see accompanying article on page 28). Only then can progress be made toward improving test scores—and determining the proper role of testing in education.

Consider for a moment the factors that contribute to a student's performance on any kind of test, whether it is teacher-made or standardized. If the test is a good one—that is, both *valid* (measures what it's supposed to measure) and *reliable* (yields consistent results)—then students' scores are determined first and foremost by knowledge and skill: Does *he know the answers* to the questions on the test? Can they do the tasks that are required? We must remember this at all times: There is no magic that will enable students to succeed on a good test if they don't know the subject matter or possess the skills that the test measures (short of cheating, that is).

Let me tell you a story about the director of research in a large public school system: He was dismayed because student scores on a national standardized reading test had dropped from the year before, and he refused to release the scores—to the superintendent, to the press, to anybody. First he asked the school system's computer center to rescore the tests; there *must* be an error, he thought. There wasn't. Then he asked the computer people to check all the statistical computations. Still no error. After several weeks and clamors on all sides to see the scores, the head of the computer center met the research director in the parking lot. "Hey, buddy," said the computer specialist, "I think I've figured out how to get those test scores up."

"How? How?" asked the relieved research director.

"Teach those kids how to read."

Given that test performance is determined largely by what students know and can do, here are some other possible, if lesser, influences on test scores:

1. **Test directions.** If the directions say to select the *least* probable explanation for an event and a student selects the *most* probable explanation, then he or she has missed the question. If one of the choices for a set of mathematics questions is, "The answer cannot be determined from the information given," and if the student ignores that choice, the student's score might suffer. If students are told to answer four out of five essay questions, and they answer all five, then they have wasted time that could have been devoted to more complete answers to four questions.

2. **Test pacing.** When there is limited time to take a test, students should pace their work so they have sufficient time to get through questions they know; they shouldn't get hung up on questions they're unsure of. They should save the difficult questions to ponder at the end of the test.

Scarvia B. Anderson is senior vice-president of the Educational Testing Service.

3. **Neatness.** A brilliant essay can be dulled by a sloppy or illegible paper, so neatness counts. So does attention to detail. Matching question numbers to spaces on the answer sheet and completely erasing answers when students change their minds can be of major importance on a machine-scored test.

4. **Physical conditions.** There is, of course, a rather broad range of tolerance before performance will be adversely affected by physical condition at the test site. Relatively poor lighting, or a testing room that is a little too warm, might make the test taker uncomfortable, but neither is likely to jeopardize test performance if the student is capable of doing the tasks on the test.

5. **Emotional or physical health.** A slight cold probably would have little influence on a student's test performance, but a bad case of the flu with accompanying fever seriously would reduce the student's ability to perform test tasks. (One of my professors described studies he did during World War II that showed aptitude test scores were affected only slightly when recruits took tests right after an all-night march.)

More troublesome are emotional disturbances, the most relevant of which is called "test anxiety." We know that some people get more anxious about a test than others; we also know that anxiety varies with how well prepared students think they are and with what they think is riding on the outcome. For example, we would expect students to be more anxious about a test that was to be used in determining their eligibility for a significant award than on a test that was only one of several in a school course. We shouldn't forget, however, that there *are* a few among us who actually like taking tests and are inclined to give our all in a testing situation regardless of the purposes of the test.

The measurement and study of test anxiety are complex and have challenged a number of psychologists through the years. In spite of all the work, we still cannot predict the number of people in a given population who will have paralyzing test anxiety; the number probably is small. On the other hand, we know that a little anxiety probably is *facilitating* rather than debilitating. (Think about your own performance when you know you are being judged by important audiences.) Even though I don't believe test anxiety is a serious determinant of performance for many people, I would not dismiss its value as an ego-saving device for thousands of students and their parents. If the term "test anxiety" did not exist, we probably would have to invent it to preserve the self-concepts of those who otherwise would have to blame their own ignorance or laziness for poor test performance.

Now that we've examined some of the major influences on test scores, we must consider what we as educators and teachers can do to reduce the negative influences.

- We certainly can design testing conditions in which directions are as clear as possible, test administrators are friendly and helpful, the physical environment is comfortable, and students who are ill are offered make up exams.

> No magic formula will enable kids to succeed on a good test if they don't know the subject matter or possess the skills that the test measures

- We also can teach students test-taking skills—pacing, handling machine-scored answer sheets, understanding different types of questions, and so on. Familiarity with objective tests is not the problem it once was. In the *first* testing study I ever did (in 1951, at George Peabody College for Teachers), we found that many of the freshmen had no previous experience with objective tests and machine-scored answer sheets. So, we instituted practice tests to try to even out any differences associated with the mechanics of test taking. Today, almost all children are exposed to a variety of standardized tests as they go through school.
- If we can identify the most extremely anxious students, we might be able to recommend personal therapy for them.

What can we do, though, about the biggest factor that affects student scores: knowledge and skills? The best answer to that is to teach reading, writing, and whatever skills tests demand. Unfortunately, if the knowledge or skills are significant, this takes a rather long time. Everyone wants to know if there are any shortcuts.

Enter coaching, one of the most confusing elements of the debate on standardized testing. The N.E.A. and others say that tests are no good if they are coachable. But this denies the link between education and the testing process. I think there is little place in education for tests on which performance *cannot* be influenced by instruction. But to discuss coaching intelligently, we first need to consider the nature of the coaching, the length of the coaching, and the kind of test that is being coached for.

Coaching usually can be distinguished from regular instruction by its duration, its purposes, and its techniques:

- *Instruction* generally is during a long-term period; *coaching* generally involves a few hours, days, or weeks.
- The purpose of instruction is to develop skills, knowledge, and understanding to an appropriate level of mastery—for their own sake. If students do better on tests as a result, fine. The purpose of *coaching*, however, is to improve performance on a specific examination. If students improve their skills, knowledge, and understanding as a result, fine.
- A good instructional program generally includes alternative explanations of phenomena and solutions to problems, allows for practice with variations, and provides a certain amount of enriching detail—all of which will help the student to understand as well as to know. Coaching, on the other hand, concentrates only on the kinds of questions that are likely to be asked on the test, on quick solutions to typical problems, and on associations of test language with certain types of responses. Drill is a significant part of most coaching.

I was coached in the evenings for three weeks for my Ph.D. examination in German. Because the examination only required me to translate passages, and because we were allowed the use of a dictionary, I did not spend much time learning long lists of vocabulary words. I *did* memorize the articles and common prepositions, conjunctions, adjectives, and adverbs so I wouldn't have to spend time looking them up. My tutor taught me how to look at a German sentence, quickly pick out the subject and predicate, and determine which words went with each. And because we knew the exam would be in my field (psychology), we practiced on abstracts of articles from German psychological journals, not on travel books or short stories. If my object had been to learn to read scientific German fluently so I could keep up with current literature, my tutor obviously would have taken a very different approach; and if I really had wanted to learn German, even more time would have been required.

I passed the exam handily, but today I know practically no German. Unfortunately, retention of facts learned during coaching frequently is poor. Coaching seems to work best when it is of the "refresher" variety: If a student has been out of school for a few years or hasn't studied a certain subject recently, coaching can clear away some of the cobwebs from skills the student once had.

The amount of time students spend being coached seems to make a difference in test performance, but only up to a certain point

The amount of time students spend being coached seems to make a difference in test performance, but only up to a point. Studies of the S.A.T. show a positive relationship between the amount of time spent in coaching and score increases. The law of diminishing returns, however, soon begins to operate. For example, an increase of 10 points on the S.A.T.'s 200 to 800-point verbal scale might be associated with 12 hours of coaching; to increase scores by 20 points, however, might require 57 hours; 30 points might require 260 hours. Extrapolating beyond existing data, it appears that for a 40-point gain on the verbal portion of the S.A.T., something close to 1,200 hours of coaching time would be required. That's full-time schooling.

Another problem: The broader and more general a test is, the more difficult is the task of preparing for it. That's what makes coaching for *aptitude* tests so difficult. It would be impossible in a couple of weeks to teach all the words that might be included on the vocabulary section of the S.A.T. But if the test were on American literature, you certainly could hit some of the high spots in two weeks. And if the test were even more sharply focused (testing knowledge of the symbols for chemical elements, perhaps), students probably could learn everything they needed to know in less than two weeks.

Here's how I would summarize my position on coaching:

- Those who would condemn tests because they are coachable ignore a fundamental relationship between testing and teaching.
- The success of coaching depends on many factors, including student motivation, the nature and duration of coaching, and the type of test. In general, aptitude tests are harder to coach for than are achievement tests.
- Whether coaching actually improves test scores, many students *believe* coaching is helpful (often, they say, because it reduces anxiety). Therefore, schools would be well advised to provide interested students with appropriate opportunities to prepare for tests—especially in test-taking skills. Just two or three sessions should do; this will help avoid charges of inequity between students who can afford expensive commercial coaching courses and those who can't.

Tests are the latest scapegoat for failures attributed to education, but the label is not likely to stick. As American Federation of Teachers President Albert Shanker put it, "You can't blame the thermometer for the patient's fever." Considering the growing public and legislative demand for educational accountability and productivity, tests probably will occupy a more, rather than a less, important place in the education scene. Consequently, it behooves educators to learn as much about measurement as they can in order to monitor testing practices, to resist inappropriate uses of tests in the school systems, to counter the misuse of test results in unwarranted attacks on education, and, most important, to harness important testing tools for their own decision-making needs. ☐

Take this

Tests are an easy mark for education's critics, partly because tests are not well understood—by educators *or* the public. Here's a crash course on the various types of tests and what they are designed to accomplish: You can increase your understanding of almost all types of education tests by determining (1) whether they measure aptitude or achievement, (2) what kind of response the student makes, (3) how the results are interpreted, and (4) who makes them up.

1. *Aptitude tests versus achievement tests.* There is a difference between aptitude and achievement tests, but it is not always immediately apparent. Performance on both types of tests depends on what students have learned, not just what they came into this world with. (A wolf child would not do well on either kind of test.) For example, Sample 1 contains two questions: Question A is from a well-known measure of scholastic aptitude; Question B is from a basic skills achievement test. In general, so-called aptitude tests measure skills that have developed over a long period of time; they also measure applications of skills in new situations. It would be unusual to look at an aptitude test question and be able to say precisely, "I learned that on Tuesday" or "Mrs. Parker taught me that." Aptitude tests are valued primarily for their *predictive* value. We say that scores on aptitude tests provide some indication of how well people will do in college, at the police academy, or in advanced math. Because the skills measured on aptitude tests develop over a long period of time and from a variety of influences, it is not fair to blame the schools *entirely* if the students score poorly on aptitude tests. Unfortunately, schools *were* blamed in many quarters when S.A.T. score declines became big news. (Of course, schools shoulder part of the blame if, after 11 or 12 years of education, students cannot handle words and number concepts as well as students of a few years before.)

Achievement tests, on the other hand, generally are more easily understood and appreciated by the public than are aptitude tests. That's because the questions on achievement tests can be related more readily to school curriculum. Scores on science, history, literature, and mathematics tests should be influenced by the quality and recency of rele-

crash course on test design

vant instruction as well as by the student's inclination to learn. It is fair to blame the school and the teachers—as well as the students—if students do badly on a test that covers material they just have completed or on a test that measures skills they have been taught for several years.

To oversimplify, the information from aptitude tests is "looking ahead" information: Do students have the knowledge or skills necessary to help them succeed in the future? The information from achievement tests is "looking back" information: Did the students learn what was taught? Of course, if students are not learning what the schools are teaching, and schools find this out in a reasonable time, they can institute procedures to remedy the situation. Consequently, one hears fewer charges of test bias leveled against achievement tests than against aptitude tests.

2. *Type of student response.* Responses basically are of two kinds. The student must recognize the appropriate answer to a question or problem, or the student must produce the answer. Recognition questions on most professionally produced tests are multiple-choice items. Teachers frequently use true/false or matching items as well. In spite of popular opinion, multiple-choice questions *can* provide an effective measure of reasoning processes as well as straight memory. Look at Sample 2, Question C. The question does not simply ask, "What areas of the world have the highest population densities?" Rather, it presents a novel situation in which the student must infer that, of the choices offered, *only population density* (Choice A) is plausible.

The most frequently used production tests are short-answer or "fill-in-the-blank" tests and essay questions. An example of a short-answer test item: "The sum of scores divided by the number of scores is called the ____________________." The student could write in "mean," "arithmetic mean," or "average." You probably are familiar with essay questions from your own school days. For example: "What is meant by the statement that France, before 1789, was centralized but not unified?" Or "Why did the Dreyfus affair become an issue of international significance?" Or, more typically, "Describe your summer vacation."

Of course, the better constructed the questions and the clearer the directions to students, the easier essay tests are to score. But even then, scoring is not simple. To avoid biased and idiosyncratic scoring, scorers need to have *in advance* a clear outline of just what components will be scored and how much weight will be given to each. Will essays, for example, be scored for style and mechanics as well as substance? A successful scoring technique used by professional test makers (but rarely by teachers) is called "holistic scoring." Raters agree on the characteristics that a top paper will have and choose samples that epitomize that standard; they do the same for papers at each of the lower ranks. Scores then are assigned to all papers by comparing them to the samples.

Production tests do have their problems: They can be tough to score; fewer such questions usually can be presented in a given period of time, so less comprehensive coverage of an area is possible. So why do testers use such questions at all? Because production tests measure characteristics that recognition tests cannot. If you want to find out how well students write, there's no substitute for asking them to write. To

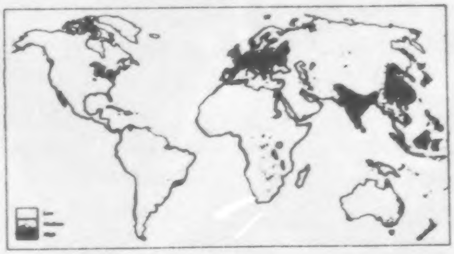
Samples of major types of test questions

Sample 1

Question A. (From an aptitude test)
In the currency of the country of Ug, 15 dops are equal to 1 tif. If 10 dops equal 1 decadop, what is the value in tifs of 6 decadops?
(A) 1/15 (B) 2/5 (C) 3/2 (D) 3 (E) 4

Question B. (From an achievement test)
If your local sales tax is 3%, what is the total cost of a $2.70 lunch?
(A) $2.73 (B) $2.78 (C) $3.00 (D) $3.51

Sample 2



Question C.
The shading on the above map is used to indicate:
- **(A) population density**
- **(B) percentage of total labor force in agriculture**
- **(C) per capita income**
- **(D) death rate per thousand of population**

Sample test questions and map reprinted by permission of Educational Testing Service, the copyright owner.

ask them questions *about* writing is not the same thing. You can think of many other examples: translating a passage from Cicero, speaking extemporaneously on a popular topic, analyzing a blood sample, editing a technical manuscript, making a soufflé, disassembling a rifle.

This distinction between recognition and production tests is an important one. A student who might recognize appropriate laboratory procedures from word and picture descriptions, for example, might well add water to sulphuric acid, instead of vice versa, when left alone in the lab.

3. *Test interpretation*. Tests can be interpreted in two ways: by reference to the performance of other people (norm-referenced) and by reference to judgmental standards (criterion-referenced). In a norm-referenced test, for example, we would say that Betty's score is in the top quarter of scores of students in her graduating class, or above the mean for a national group of twelfth graders, or at the 60th percentile for seniors in the state (meaning that 60 percent of the state's seniors scored *lower* than she did). This example also indicates that, with respect to the skills measured, Betty had more competition in the state than in her own class (note that she stood higher in her class than she did in the state as a whole).

Criterion-referenced tests are interpreted according to judgments about what constitutes appropriate performance—highly satisfactory, adequate, or minimal—depending on the test and the purpose for which it was given. For example, a group of teachers might agree that a score of 60 was a passing score on Test X or that ten correct answers out of 12 items on a math quiz indicated acceptable mastery of a concept. An important thing to keep in mind is that there are no universal criteria for such tests: Sixty percent might be a passing score for one test but not passing on an easier test; ten correct answers out of 12 items might be considered mastery for one concept but not for another.

It is unfortunate that criterion-referenced and norm-referenced measurement often are viewed in sharp contrast, sometimes with the attitude that using one is "good" practice and the other "bad" practice. In fact, the two approaches frequently overlap and certainly are useful supplements to each other. Moreover, normative considerations usually underlie the choice of standards for criterion-referenced tests (although these considerations might be implicit rather than explicit). For example, we may set 50 words per minute with two errors as the minimum standard for graduation from a three-month typing course. We don't set ten words per minute or 150 words per minute because these rates do not reflect *typical* typing requirements or *average* human capability—both normative notions.

Whether test users rely more heavily on norm-referenced or on criterion-referenced interpretations of test performance, they still must rely to a great extent on *judgment* in setting standards or cutoff scores. In a school system experiencing a teacher surplus, for example, the accrediting agency might qualify only those teacher applicants who score above the 25th percentile on a test. In times of a teacher shortage the school system might lower the standard.

A mysterious phenomenon is that many teachers and parents take the 50th percentile, or the mean, as the implicit standard of adequate performance. They will bewail the fact that little Maria or Fred is below average without stopping to think that half the children, by definition, *have* to be. In fact, it is a mistake to think of "average" as representing a single score; it is better to think of "average" as a range of scores encompassing the middle third of the group.

There are many different procedures that educators use to set standards or cutoff scores. These vary in formality and elaborateness. The burden is on educators, however, to make the standards they adopt explicit and to provide a clear rationale for them. This is of paramount importance if cutoff scores are to be used for such crucial decisions as graduation, promotion, or admission to a program of study.

4. *Test developers*. By far the largest number of tests are those developed by individual teachers. The critics of testing usually overlook this fact. For every S.A.T., there are thousands of midterm and final exams in mathematics and English; for every standardized elementary reading test, thousands of dittoed reading exercises are used in classrooms. Classroom tests (and other evaluations) are, and will continue to be, the major factors in promotion, graduation, recommendations for further education and employment, and students' self-appraisals of their academic fitness.

The two major types of outside groups that develop and supply tests are state departments of education and national testing agencies (Educational Testing Service; American College Testing Program; Psychological Corporation; Harcourt, Brace, Jovanovich; Houghton-Mifflin; CTB/McGraw-Hill). Sometimes, states and national agencies work together, especially when state departments of education have limited technical or technological resources.

In the case of achievement tests, those developed and used to evaluate students at the local level and those selected or imposed from outside sources should be complementary. A mandated state basic skills test should not come as a shock to a student. If the state test is an appropriate measure of school learning, it should look like the exercises students already have seen—frequently—in their schools. A lack of congruency is a legitimate concern of education critics: Either schools have not prepared students properly, or the outside test is an unfair reflection of school objectives.

In the case of aptitude tests, you would not expect tests developed by outside sources to bear as close a relationship to the tests normally used in schools. (Teachers simply don't make up aptitude tests.) This is not a fault: For example, a student might encounter on an aptitude test a vocabulary word that is unfamiliar. What we *hope* is that from an accumulated knowledge of words and how they are used, the student might be able to deduce the meaning from the structure of the word and from the context.

By now, you should be suitably impressed by the number and variety of tests that can be devised from the four classifications: There are *criterion-referenced achievement* tests that are developed by *teachers* and require students to *produce* the answers (this category probably characterizes the largest number of educational tests given and taken in the world). There are *criterion-referenced, multiple-choice* tests that are developed by *state departments of education* to measure *achievement* in the basic skills. There are *norm-referenced, multiple-choice, aptitude* tests such as the ones designed by the *Educational Testing Service* for the College Board. (Tests of this type have been of greatest concern to school people in recent years, probably because they are the ones over which school people feel they have the least direct control.) In the appropriate circumstances, each type of test can be valuable. —*Scarvia B. Anderson*

Attachment C

DEVELOPING A TEST

EDUCATIONAL TESTING SERVICE

DEVELOPING A TEST

From the time a standardized test is conceived to the time it is administered, dozens of people are involved: test specialists, editors, reviewers, teachers, and professors in the knowledge or skill to be measured. At Educational Testing Service, developing an entirely new test takes about two years, and it takes about 18 months for each question to pass through all its development stages.

Structuring examinations and writing questions are extraordinarily complex, but not mysterious. To help in understanding the process, here are its highlights.

1. Defining Educational Objectives

The decision to construct a test of particular aptitudes or achievements usually results from the requests of educators who see the need for such a test. Once the decision is made, test developers ask:

- Who will take the test and for what purposes?
- What can test takers be expected to know about the field?
- How should test takers be able to use their knowledge?
- What kinds of questions should be in the test and how many of each kind?
- How long should the test be?
- How difficult should the test be?

2. Advisory Committees

At ETS, these questions are usually answered with the help of expert advisors or examiners from all across the country.

Each testing program has its own committee of advisors. For example, the National Teacher Examinations are being revised by a panel of 25 leaders in school administration and teacher education working with ETS staff.

3. Question Writing

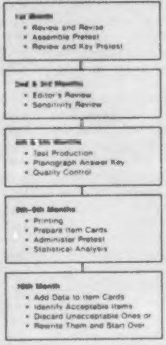
Once specifications have been established, decisions are made on the types of questions to use. For standardized tests, ETS most frequently uses the objective, multiple-choice format, although some tests also include essay questions.

Advisory committees, outside experts, and ETS staff write the questions.

An Example

The Cycle Begins:

Write the Items


1st Month
• Review and Revise
• Assemble Pretest
• Review and Key Pretest
2nd & 3rd Months
• Editor's Review
• Sensitivity Review
4th & 5th Months
• Test Production
• Planograph Answer Key
• Quality Control
6th–9th Months
• Printing
• Prepare Item Cards
• Administer Pretest
• Statistical Analysis
10th Month
• Add Data to Item Cards
• Identify Acceptable Items
• Discard Unacceptable Ones or
• Rewrite Them and Start Over

4. The Pretest

Once questions are written, they are administered to a sample group similar to the people for whom the test is intended. The *pretest* tells test developers several things:

- The difficulty of each question.
- Whether individual questions are ambiguous or misleading.
- Whether items should be revised or discarded.
- Whether incorrect alternative answers should be replaced or revised.

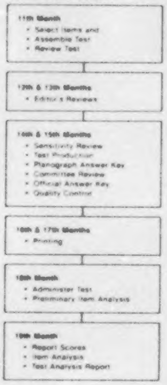
5. Sensitivity Review

Before and after the pretest, each test is reviewed to ensure that the questions reflect the multicultural nature of our society and that appropriate, positive references are made to minorities and women. Each test item is reviewed to ensure that any word, phrase, or description that may be regarded as biased, sexist, or racist is removed.

An Example

The Cycle Continues

Assemble the Test



These judgments are made by "sensitivity" reviewers who are especially trained to identify and eliminate material that might be unfair or offensive to any group.

6. The Final Phase

In the final phase, test makers analyze the pretest data and choose appropriate questions of the proper difficulty to reflect the subject matter and to test specific skills. A proper distribution of content accounts for differences in school curricula as well as regional differences, rural-urban differences, and sex and ethnic differences.

After the test is assembled, it is reviewed by other specialists, committee members, and, sometimes, by outside experts. Each reviewer answers all questions independently and prepares a list of correct answers which is then compared with the intended answers to verify agreement on the correct answer to each question. **No test can go to press until the person responsible certifies that at least three different people have agreed to the correct answers to every question.**

The test is then sent to the external committee of examiners for review before going to press.

7. Test Validity and Disclosure

The test preparation process continues even after the test has been administered. Continuing research is necessary to ensure that the test is valid—that it does the job it is designed to do.

After the test is administered, but before final scoring takes place, preliminary statistical analysis of each question is carried out, based on several thousand answer sheets. The results are reviewed question-by-question. If a problem is found, corrective action, such as not scoring the question, is taken before final scoring and score reporting.

Accuracy of test content is further strengthened by feedback from students to the test center supervisor, alerting test administrators to the possible existence of ambiguously worded items. This process has been extended through test disclosure, which enables students to examine the questions and answers after the test has been given. When test development errors come to light through these processes, appropriate corrective action is taken.

• • •

At every stage of the process, ETS does all that is possible to assure the accuracy and fairness of individual test questions and the examinations as a whole. It's a demanding and time-consuming process . . . but it's worth the effort. ETS strives to set high standards in the field for quality tests.

For further information, contact Information Division, ETS, Princeton, NJ 08541.

007401 • S71P1 • Printed in U.S.A.

Attachment D

QUESTIONING THE QUESTIONS
The Process for Challenging a Question on an ETS Test

1. Test-Taker Inquiry to ETS

Test-takers may inquire about a question on an ETS admissions test in several ways:

(1) Speak to the supervisor at the testing center, who will file a report to ETS concerning the test question in dispute

(2) Write or call ETS immediately after the test explaining the difficulty encountered with the particular question

(3) Write or call ETS about a disputed question after reviewing the test questions provided the test-taker following a disclosed test administration

2. ETS Staff Review

Each inquiry is reviewed by two or more ETS staff members who have major responsibility for the type of question in dispute. Additional reviews by subject matter specialists, such as university faculty in the relevant field, are also frequently obtained -- depending on the nature of the challenge.

3. ETS Response to Test-Taker

A letter, either explaining the question's soundness, or acknowledging a flaw, is then mailed to the test-taker. (If a flaw is acknowledged, the next step in the process is #8 below.)

4. Further Test-Taker Inquiry

If the test-taker is not satisfied with ETS' response, he or she may write ETS again to continue the challenge.

5. Additional Independent Review

Additional reviews of the disputed test question are obtained from a combination of university faculty and/or other appropriate experts outside ETS (such as secondary school teachers for the PSAT or SAT) and senior ETS staff who were not involved in developing the test in question. Without prior knowledge of the test question or correspondence related to it, the independent reviewers are asked to select an answer to the question and justify their choice. They are then given the correspondence between the test-taker and ETS and asked to explain in writing why their perceptions are or are not changed by those arguments.

6. Second Response to Test-Taker

A second letter is then sent to the test-taker, either explaining the arguments in defense of the challenged question, or acknowledging a flaw. (To date, no test-taker has continued a challenge beyond ETS' second response.)

7. Appeal Process

In the event that a test-taker should remain unsatisfied with ETS' second response, an appeal process has been included in the procedures for handling test question inquiries adopted jointly by the College Board and ETS relative to College Board programs. (ETS will recommend to other test sponsors comparable procedures should the need arise.)

The College Board-ETS appeal process calls for a judgment based on all materials related to the case by a three-member external appeal panel chosen by the chair of the appropriate College Board Academic Advisory Committee (i.e., Mathematics, English, Science, etc.)

8. Possible Outcomes

If a test-taker's challenge is substantiated at any point before scores are reported, corrective action is taken immediately, typically, by deleting the disputed question prior to scoring the answer sheets.

If a flawed question is discovered after scores have been reported, a determination is made by the test sponsor, with consultation from ETS, relative to rescoring answer sheets and rereporting scores. This decision may vary according to the testing program. However, considerations affecting the decision include fairness to students, the accuracy of measurement, how the test is used, and the extent of impact rescoring would have on test-takers remaining in the admissions decision-making process. Flawed questions are withdrawn from further use.

9. Errors are Rare

ETS, like other organizations, is not perfect. Although mistakes are rare, flawed questions or mistakes sometimes slip through the rigorous, 18-month process that is used in developing and refining tests. [Out of 5,695 questions on disclosed tests from Jan. 1980 to June 1981, only three were found to be flawed, less than 1/10 of 1%.] Whenever an error is discovered, corrective action is taken promptly and in accordance with the guidelines stated above.

Prepared by The Educational Testing Service, July, 1981.

INQUIRIES ABOUT TEST QUESTIONS FROM TEST-TAKERS
(Includes All Administrations)
January 1, 1980 - June 30, 1981

This table points out that:

- The large majority of inquiries about test questions for all six programs is made by test-takers shortly after the administration of a test, and prior to the receipt of test questions provided by test disclosure.
- a very small portion of the 1980 and 1981 inquiries were received as a result of test disclosure.

	GMAT	GRE Aptitude	LSAT	SAT	TOEFL	PSAT	TOTALS
(1) Reports by Test Center Supervisors	89	367[a]	118[b]	107	2	11	695
(2) Direct Contacts to ETS Before Scores Reported	11	20	11	15	3	1	61
(3) Direct Contacts with ETS After Receiving Test Questions & Answers	6	6	4	6	1	9[c]	32
Total Inquiries	106	393	136	128	6	21	788
Faulty Questions Identified by:							
(1)	0	1	1	3	2	0	7[d]
(2)	0	0	0	0	0	0	0
(3)	0	0	1	1	0	1	3
Total	0	1	2	4	2	1	10

Notes:

[a] Large number due to multiple comments on a few questions.

[b] Almost half of these were on one question which contained a misprint; the question was not scored.

[c] Includes letters from parents and school personnel.

[d] Flawed questions identified before scores are reported are not counted in scoring.

NUMBER OF QUESTIONS DISCLOSED & NUMBER OF QUESTIONS CHALLENGED SUBSEQUENT TO DISCLOSURE OF TESTS ADMINISTERED BY ETS JANUARY 1980 - JUNE 1981

Test	No. of Tests Disclosed	No. of Questions Disclosed	No. of Questions Challenged	No. of Flawed Questions	Disposition Made
GMAT	6	900	4	0	
GRE Aptitude	5	1,025	6	0	
LSAT	7	1,330	4	1	Rescored 23,000 answer sheets; revised 22,000 scores.
PSAT	2	230	9	1	Rescored 830,000 answer sheets; revised 250,000 scores.
SAT	8	1,160	6	1	Rescored 85,000 answer sheets; revised 17,000 scores.
TOEFL	7	1,050	1	0	
TOTAL	35	5,695	30	3	Rescored 938,000 answer sheets; revised 289,000 scores.

Attachment E

STATE LEGISLATURES IN WHICH BILLS TO REGULATE AND DISCLOSE ADMISSIONS OR OCCUPATIONAL TESTS HAVE BEEN INTRODUCED

States in which bills were introduced prior to 1981	States in which bills were introduced prior to 1981 but which did not consider the issue in 1981	States in which bills were introduced for first time in 1981	States in which legislation requiring disclosure of test content has passed
California	Colorado	Alaska	New York
Colorado	Indiana	Arizona	
Connecticut	Maryland	Georgia	
Florida	Mississippi	Oregon	
Illinois	Ohio	South Dakota	
Indiana	Rhode Island		
Louisiana	South Carolina		
Maryland	Tennessee		
Massachusetts	Texas		
Minnesota	Washington		
Mississippi			
Missouri			
New Jersey			
New York			
Ohio			
Oklahoma			
Pennsylvania			
Rhode Island			
South Carolina			
Tennessee			
Texas			
Washington			
22	10	5	1

TEACHERS COLLEGE COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027

DIVISION OF PHILOSOPHY
THE SOCIAL SCIENCES, AND EDUCATION

July 13, 1981

Congressman Paul Simon
Subcommittee on Postsecondary Education
320 Cannon House Office Building
Washington, D.C. 20515

Dear Congressman Simon:

I am writing to register with you my strong opposition to a national "truth-in-testing" law. I am a historian of education at Teachers College, Columbia University; my overriding interest as an educator is in the quality of American education. I see no valid legislative purpose to be served in the enactment of federal legislation regulating educational testing. Such a law would be an intrusion into the educational process for spurious purposes and would introduce a federal presence where none is needed.

I base my opposition to such a law on the following facts:

1) No college or university is compelled to use a standardized admission test; they ask prospective students to take such tests because they want to know how they perform on a test of verbal and mathematical skills and how they compare to others of their age who take the same test; institutions ask students to take these tests, because they feel that the tests are a useful predictor of future performance, thus enabling the college to determine whether a student has the requisite skills to succeed and to place entering students in appropriate programs. If postsecondary institutions do not want this information or do not trust it, they do not ask for it.

2) Test scores serve a useful purpose in comparing students from different secondary schools and different backgrounds. Admissions officials find that they need some objective measure by which to compare the "all-A" student from a well-known high school and the "all-A" student from an unknown high school; because of the well-established phenomena of grade inflation, college admissions officers can no longer be sure that a student with high marks is actually prepared to succeed in an academically demanding college program.

3) Few students fail to get into the college of their choice, and everyone who wants to go to some college can get in, without regard to test scores. The College Board polled its members recently and discovered that only 8 percent of the nation's postsecondary

institutions were competitive, in the sense of having more applicants than places; 34 percent admitted everyone who applied, regardless of their previous academic experience; and the remaining 58 percent accepted everyone who met their requirements (such as, having a high school diploma or a high school average of 80). In the last-named group of institutions, everyone or almost everyone who applied was admitted.

4) In the face of declining numbers of college-age students, the traditional admissions "game" has virtually been reversed. With the exception of a handful of elite institutions, colleges are now in the position of competing for students, instead of students competing to get into college.

5) Test scores, in short, keep no one out of college, but do provide a useful piece of information that is helpful in matching students and institutions and in diagnosing student strengths and weaknesses.

6) One of the hidden purposes of "truth-in-testing" legislation is to disrupt the process that makes standardization of test questions possible. The reason that a standardized test is reliable is because the questions on it are used and reused, which makes it possible to know whether the test item contains any cultural bias or whether it is flawed in ways not immediately apparent; a substantial part of present tests consists of questions that are being tried out, without counting on the test-taker's score. If test items were disclosed every time a test were administered, it would make the process of test-construction more expensive and less reliable.

Most important, it seems to me, is the question of why federal legislation should be written which would interfere with the standards and criteria of postsecondary institutions. Behind the "truth-in-testing" drive, I believe, are well-meaning people who would like to ban standardized testing altogether. What they would really like to ban is the competition to get into a few highly select colleges and universities. Common sense suggests, however, that this competition will exist whether it is based on standardized testing or or individual tests administered by each institution that has more applicants than places available. Common sense also suggests that this is not an appropriate arena for federal legislation.

Yours truly,

Forgetting the Questions

The Problem of Educational Reform

DIANE RAVITCH

IT WOULD BE DIFFICULT TO FIND a sustained period of time in our history when Americans felt satisfied with the achievements of their schools. From the early nineteenth century on, it has been commonplace to find a fairly consistent recitation of complaints about the low state of learning, the poor training of teachers, the insufficient funding of education, the inadequacies of school buildings, and the apathy of the public. The temptation exists to attribute the concerns of the 1980s to this strain of despair about the historic gap between aspiration and reality, this sense that schools have always and will always fall short of their mission. But it would be wrong to do so, not only because it would encourage unwarranted complacency, but because the educational problems of the present are fundamentally different from those of the past.

One important difference is that so much of the past agenda of educational reformers has been largely fulfilled. In one sense, the educational enterprise is the victim of its own successes, since new problems have arisen from the long-sought solutions to earlier problems. Idealistic reformers, eager to improve the schools and to extend their promise to all children, sought the appropriate lever of change. *If only* teachers had college degrees and pedagogical training; *if only* teachers would band together to form a powerful teachers' union; *if only* there were federal aid to schools; *if only* all children were admitted to school regardless of race or national origin; *if only* all students of high ability were admitted to college; *if only* colleges could accommodate everyone who wanted to attend; *if only* students had more choices and fewer requirements in their course work; *if only* schools were open to educational experimentation; *if only* there were a federal department of education. . . . The "if only" list could be extended, but the point should be clear by now. All these "if onlies" have been put into effect, some entirely and others at

---

DIANE RAVITCH, associate professor of history and education at Teachers College, Columbia University, is the author of *The Great School Wars: New York City, 1805–1973* and *The Revisionists Revised: A Critique of the Radical Attack on the Schools*. She is currently at work on a history of American education since 1945.

least partially, and rarely have the results been equal to the hopes invested.

In reality, many present complaints are reactions to hard-won reforms of the past. Though the educational preparation of teachers is more extensive than ever, at least when measured by degrees and years of formal schooling, the education of teachers is still a subject of intense criticism. The realization has dawned in many quarters that a credential from a state university or a school of education is no guarantee that its bearer knows how to teach or what to teach, loves teaching or loves learning. Nor are today's critics delighted by the undeniable power of teachers' unions. True, the unions have used their political clout to improve teachers' salaries and to win vastly enlarged federal education expenditures, but unionization has not produced the educational changes that some of its advocates had anticipated. Similarly, the sense of achievement that should have followed the removal of racial barriers to higher education quickly gave way to concerns about social stratification, vocationalization, and declining quality. The reforms of the 1960s were effective, though not in the way that reformers had hoped. Now everyone who wants to go to college can go to *some* college, though not necessarily that of his first choice. By 1980, at least one-third of all institutions of higher education admitted everyone who applied, more than one-half accepted most or all of those who met their qualifications, and less than 10 percent were "competitive," that is, accepted only a portion of qualified applicants. As college enrollments decline, the number of competitive colleges will grow fewer. Curricular reforms have broken down the coherence of the liberal arts curriculum, both in high school and college, so that students have a wide degree of choice and few requirements. And a federal department of education has at last been established, though with what benefits or burdens for schools and children it is too soon to say.

Yet having won so many victories, some of truly historic dimension, American education is still embattled, still struggling to win public support and approval, and, perhaps worse, still struggling to find its own clear sense of purpose. Paradoxically, the achievements of the recent past seem to have exhausted the usually ready stock of prescriptions for school reform and to have raised once again the most basic questions of educational purpose.

Like other major institutions in our society, the schools are continually judged by today's demands and today's performance, and no credit is extended by clients or critics for yesterday's victories. Which is as it should be. School criticism, as I noted earlier, is nothing new. Behind any criticism, however, are assumptions about what schools should do and can do, and criticisms have shifted as assumptions about the goals and potentialities of schools have changed. Since the early nineteenth century, the tenor of school criticism has been essentially optimistic; no matter how despairing the critic, his working assumption has been that schools are valuable institutions, that they have within them the power to facilitate great social, moral, and political regeneration, and that more money, or more public concern, or better teachers could extend the promise of schooling to everyone. If more people had more schooling, critics have contended, and if schools were amply financed and well staffed, there would be enormous benefit to the individual, the society, the economy, and the body politic. With relatively little dissent, Americans have believed in schooling—not because of a love of the hickory stick and the three Rs, or (as some latter-day critics would have it) because of the schools' ability to make children docile workers, but because Americans are deeply committed to self-improvement and the school is an institutionalized expression of that commitment.

Participation in formal schooling has grown sharply in recent decades. The proportion of seventeen-year-olds who graduated from high school grew from about 50 percent in 1940 to about 85 percent in the late 1970s. Similarly, the proportion of young people who entered college climbed from about 16 percent in 1940 to about 45 percent in 1968, at which time it leveled off. In no other country in the world does participation in formal schooling last as long, for so many people, as in the United States. To understand why this broad democratization of educational participation occurred, as well as why the 1980s began on a note of disillusionment, it is useful to consider some of the expectations we have attached to formal schooling.

Until well into the twentieth century, only a small minority of Americans attended college. College was not only expensive but exclusive. Many, perhaps most, colleges maintained quotas for some groups (like Jews and Catholics) and excluded others altogether (blacks). After World War II, more than two million veterans attended college, crowding and sometimes overwhelming America's campuses. The GI Bill launched the world's first experiment in universal access to higher education. While most veterans did not use their benefits to attend college, the experience of those who did benefited the individuals, the institutions, and the economy. In light of the success and popularity of the GI program, the conviction that college should be a right rather than a privilege gained broad support.

While demands for expanded access to higher education grew steadily in the states and nation, other political forces combined to advance the role of education as a weapon against poverty. The notion that knowledge is power was certainly not novel, nor was the very American belief that schooling is an antidote to crime, poverty, and vice. The school promoters of the early nineteenth century repeatedly argued that schooling would give people the means to improve themselves and thereby break the cycle of poverty. During the early 1960s, this traditional rhetoric was given new life by scholars and policymakers. Educational programs burgeoned as an integral part of the federal government's war on poverty. Jacob Riis had written in 1892, "the more kindergartens, the fewer prisons"; in 1965 Lyndon Johnson predicted that the lives of children in the Head Start summer program would be spent "productively and rewardingly, rather than wasted in tax-supported institutions or in welfare-supported lethargy." The hope of eliminating poverty and inequality provided the major rationale not only for Operation Head Start but for general federal aid to education as well.

By the time the period of educational expansion reached a high tide in the middle 1960s, much was expected by a variety of publics. It was hoped that more education would:

- Reduce inequality among individuals and groups by eliminating illiteracy and cultural deprivation.
- Improve the economy and economic opportunity by raising the nation's supply of intelligence and skill.
- Spread capacity for personal fulfillment by developing talents, skills, and creative energies.
- Prove to be an uplifting and civilizing influence in the nation's cultural life by broadly diffusing the fruits of liberal education.
- Reduce alienation and mistrust while building a new sense of community among people of similar education and similar values.
- Reduce prejudice and misunderstanding by fostering contact among diverse groups.
- Improve the quality of civic and political life.

These hopes and expectations were a heavy burden for the schools to bear. Perhaps predictably, they did not accomplish all that was asked of them. Most of the problems that were laid at the schools' doors remained just as problematic years later (and some critics would argue that the provision of more schooling had produced the opposite effect in every instance). Poverty and inequality did not cease; their roots were elsewhere, and the schools were not able to cure deep-seated social and economic ills. While the disadvantaged received more schooling, so did the advantaged. Many poor youths entered the middle class by using educational opportunity, but others remained as poor as their parents. The value of a high school diploma declined not only because its possession became nearly universal but also, and most important, because high school graduates were not necessarily literate—mainly because of the well-intended effort to keep as many youths in school for as long as possible and to deny no one a diploma, regardless of his educational development. Society's investment in education probably did spur economic development, but it did not prevent the emergence of skepticism about the desirability of economic growth; in fact, it was precisely among the educated (and the advantaged) that economic growth became suspect because of its association with the bureaucratization, centralization, and depersonalization of modern economic life. It is impossible to gauge the effects of increased schooling on popular culture or high culture. Television, which invariably seeks the largest possible audience, undoubtedly has more power to shape popular culture than schools do (a mixed blessing, since television disperses both sitcom pap and major cultural events to mass audiences). Participation in popular culture and high culture has surely been broadened, yet it is arguable whether the quality of either has been elevated during recent decades. Nor is it possible to demonstrate that increased educational participation has eliminated distrust between groups or contributed to a new sense of community. On the contrary, educational institutions have become settings for expression of militant particularism along racial, religious, ethnic, sexual, cultural, and linguistic lines. Very likely the differences among groups have been accentuated in the past twenty years. But again it would be difficult to hold the schools directly responsible for these trends. More likely, it appears, the schools are the stage on which such issues are acted out rather than the cause of their appearance. Nor can the schools claim to have improved the quality of political life, since political participation has waned along with public regard for political institutions. But, once again, it was not the schools that were responsible for the apparent ebbing of civic commitment and the surge of political apathy, nor could they even serve as a counterforce against such attitudes. The same attitudes of distrust, skepticism, hostility, and apathy eroded the schools' own status in the social order. The same confusions that pervaded the social atmosphere also pervaded the schools. If they failed to teach citizenship, it was at least in part because teachers and parents were confused about what a good citizen was and whether "citizenship" could be taught without imposing a partisan interpretation. In short, a society that is confused and contentious cannot look to its schools to straighten things out, for the schools will reflect the same confusion and contention.

In retrospect, it was folly to have expected the schools to transform society or to mold a new kind of person. The schools are by nature limited institutions, not total institutions. They do not have full power over their students' lives (even total institutions, like prisons, have discovered the difficulty of shaping or reshaping the lives and minds of those they fully control). Schools are not fully independent in their dealings with students; they are interrelated with, and dependent on, families, churches, the media, peer groups, and other agencies of influence. Nor can schools be considered as if they were machines, all operating in the same predictable manner. Teachers vary, administrators vary, students vary, communities vary, and therefore schools vary. The schools, being complex human institutions composed of actors with different goals, different interests, and different capacities, cannot be treated as if they were all interchangeable.

As it became clear that more schooling would not provide any magical solutions, the utopian hopes once focused on the schools dissipated. Having briefly been the repository of grand and even grandiose dreams of human betterment, the schools became a scapegoat for all the wide-ranging problems they had failed to solve. Having revealed that they were but fallible instruments of social change and that any change they promoted would only be incremental, the schools became the object of rage and scorn. They were portrayed as intractable, bureaucratic, even malevolent barriers to social change. But just as it was unrealistic to believe that the schools had the power to remake society by molding those who passed through their doors, it was equally unrealistic to assert that they were powerless, meaningless, superfluous institutions with no purpose other than the care and feeding of their own employees.

Nonetheless, when the dream of a school-led social revolution faded, school criticism shifted in tone. The voices of liberal critics—those who believed that men and women of goodwill might work together to improve schools by using this program or that curriculum—diminished to mere whispers. They were drowned out by critics who believed that only radical changes in teaching or in governing schools could "save" them; by those who believed that the public schools were beyond redemption and ought to be replaced by "free" schools; and by those who advocated the abolition of compulsory schooling and the "de-schooling" of society. For a time in the late 1960s and early 1970s, bookstore shelves fairly bulged with apocalyptic predictions about the imminent demise of schooling. One book, playing on the then current phrase "God is dead," was titled *School Is Dead.* While some of the writing of this period contained sharp and telling portraits of insensitive teachers and uncaring bureaucrats, others gave vent to undisguised anti-intellectualism in their attacks on academic standards, discipline, science, and rationality. In the larger culture—and, alas, especially in academic institutions—a great revival seemed to sweep the land, casting aside "old" doctrines of deferred gratification, structured learning, and professionalism while espousing mysticism, Eastern religions, the occult, astrology, and whatever else promised to touch the spontaneous, untrained inner spirit.

These trends had curricular and programmatic consequences. In colleges, students demanded, and usually won, the abolition of course requirements, the adoption of pass-fail grading, the de-emphasis of competition and testing, and extensive choice in selecting their own programs of study. As requirements for admission to college were relaxed, high schools soon succumbed to many of the same pressures that had changed the colleges: course requirements were eased, new courses proliferated, academic standards dropped, homework diminished, and adults generally relinquished their authority to direct student learning. At all levels, both in college and high school, educational administrators reduced, to the extent possible, the schools' role as in loco parentis. To some extent, this period of student assertiveness and adult retreat was the educational side of the movement against the war in Vietnam, which provoked youthful revolt against authority in many parts of the society and the culture. But even after the war ended, there remained a lingering hostility to science, technology, and reason—as though these were the root causes of the hated war.

As the 1980s opened, it appeared that this wave of anti-intellectualism had spent itself, for complaints about the schools suggested entirely different concerns. The well-publicized decline in Scholastic Aptitude Test (SAT) scores created a context for worrying about a national deterioration in the quality of education. Not that the SAT scores were important in themselves, but they provided a sense of a pattern-in-the-carpet that had not previously been definable. For several years college officials had reported a steady increase in the number of freshmen who read poorly and wrote atrociously; the phenomenon of remedial reading and remedial writing classes spread throughout higher education, even to elite institutions. The apparent explanation, at first, was that so many new students from poor families had begun to attend college, but analysis of the SAT drop showed that the score decline continued long after the socio-economic profile of the college-going population had stabilized. Bits and pieces of evidence from other sources began to fit together. Other standardized measures of academic ability reported score declines paralleling the SAT's. National newsmagazines discovered a writing crisis and a literacy crisis. Educational malpractice suits were filed by disgruntled parents because their children had received a high school diploma in spite of being "functionally illiterate." The Council for Basic Education, a lonely voice for liberal education since its founding in 1958, found itself back in the educational mainstream, while still a lonely voice for liberal education. Demands for minimum competency tests seemed to spring up spontaneously in almost every state, though no national organization existed to promote or coordinate the movement. As concern for educational standards spread in the middle 1970s, demands for testing grew—not only minimum competency tests for high school graduation but tests at critical checkpoints in the lower grades and tests for would-be teachers. Reaction against these demands was not long in coming. The assault upon standardized testing was led by consumer activist Ralph Nader and the National Education Association. Nader released a lengthy attack on the credibility of the SAT, the most widely used college admission test, and lobbied successfully in New York State and elsewhere for passage of a "truth-in-testing" law.

While it did generate controversy, the dispute over testing was superficial, for tests were neither a cause of nor a remedy for the underlying malaise in American education. Nearly all the educational controversies of the 1970s—whether over bilingualism or sex education or testing or open admissions or busing—dealt with some aspect of the educational process that was of great importance to some constituency, but none directly raised these questions: What does it mean to be an educated person? What knowledge is of most worth? Are the graduates of our schools educated people?

The very absence of such questioning suggests a failure in educational thinking. Educators and, most especially, educational policymakers have fallen into the habit of analyzing school issues almost entirely in sociological and economic terms. In recent years it has been customary to think of schooling as a quantifiable economic good to be distributed in accordance with principles of equity or in response to political demands. The sociological-economic perspective has come to dominate educational discussion and has informed public policy. Without doubt it has contributed to necessary changes in patterns of schooling, by redirecting resources in a fair manner and by opening up access to educational opportunities. But the functionalist perspective became dysfunctional when it crowded substantive educational concerns off the policymakers' agenda, when the desire to keep students in school was unaccompanied by interest in what they would learn while they stayed in school. What I am suggesting here is not a conflict between the functionalist perspective and the educational perspective, but the danger of analyzing the schools through only one of the two prisms. There has been a fairly persistent tendency, I would argue, to neglect the role of schools as educational institutions, to treat them as sociological cookie cutters without regard to the content of their educational program. When I consider why this is so, I conclude that there are several possible explanations.

First, the sociological perspective has become dominant because it relies on quantifiable data that are accessible. It is far easier to gain information about years of educational attainment and socio-economic status than it is to ascertain the conditions of learning in any given school. Educators cannot agree on how to ascertain the educational climate or even on what should be learned. Thus it becomes irresistible to deal with, perhaps even become the captive of, data that are both available and measurable.

Second, the sociological perspective is a useful adjunct to the concept of the school as a tool of social reform. By checking on who is in school and for how many years, and on how their social background relates to their choice of occupation, we can attempt to monitor how educational resources are allocated and whether schooling is contributing to social progress. While it is neither new nor unusual to regard the school as a lever of social reform, it is unusual and perhaps unwise to see the school *solely* as a tool of social reform and *solely* as a resource to be redistributed. One consequence is that the school's diploma is confused with the learning that it is supposed to represent. In recent years, policymakers have sought to equalize educational attainment (years-of-schooling) without regard to the quality of education. This is like putting people on a diet of eighteen hundred calories a day without caring whether they are consuming junk food or nutritious food. Years-of-schooling, or a diploma, has been treated as an end in itself. Thus we have seen courts require school districts to present a diploma to students who could not meet minimum state standards of literacy, as if it was the diploma itself they needed rather than the learning that the diploma is supposed to signify. When school reformers in the nineteenth century advocated universal education as a way of improving society, they meant a broad diffusion of knowledge and wisdom, not a broad diffusion of diplomas.

Third, educational analysts have relied on the sociological perspective because it is easier to raise the level of educational attainment than it is to raise the level of educational quality. Staying in school, not dropping out, and getting a diploma represent a clear, unambiguous goal that everyone can understand without quarrel. As soon as school officials begin to define what should be taught and learned during those years, disagreements arise which are best settled by making the schools all things for all people.

For these reasons and others, educational policymakers have tended to view schooling as an instrument to achieve some other goal, only rarely as an end in itself. To the extent that they do so, they rob schooling of the very attributes that give it power. If a young man or woman has a high school diploma but can scarcely read or write, then the diploma is worthless. When a diploma, either at the high school or college level, represents a certificate of time served but not of the systematic development of intelligence and skill, then it is difficult to know why it should have any inherent value. And of course it does not.

An educational critique of schooling would have as its starting point, I believe, the idea that the essential purpose of schooling is to develop the powers of intelligence: thinking, reflecting, observing, imagining, appreciating, questioning, and judging. Beyond that, schooling has many additional purposes and serves many additional purposes, both for the student and for society. Educational literature teems with lists of the many ways in which schools should meet individual and social needs. But the schools' first purpose is to encourage and guide each person in the cultivation of intelligence and the development of talents, interests, and abilities. Schools do many other things as well; they may provide food, social services, psychological services, medical care, and career guidance. But no matter how well or how poorly they fulfill these functions, the schools must be judged in the first instance by how well they do those things that only they can do. We expect the schools to teach children command of the fundamental skills that are needed to continue learning—in particular, the ability to read, write, compute, speak, and listen. Once they have command of these skills, they should progress through a curriculum designed to enlarge their powers. Such a curriculum would contain, for every student, history and social studies, language and literature, mathematics, science, and the arts. Students need to learn these skills and disciplines in school because, except for those rare individuals who can educate themselves without a teacher, they are unlikely to have another chance to do so.

The schools are responsible both for preserving a sense of the past and for providing the ability to think about, and function in, the present and the future. More than any other educational agency, they ought to have an intelligent understanding of the inexorable connection between past, present, and future. Certainly there is disagreement about the meaning of the past and how it relates to the present and the future, and awareness of such disagreement is often invoked to justify educational aimlessness. But much of what seems to be dissension is a chimera; democratic debate ought not to be confused with chaos, nor should pluralistic politics be confused with anarchy. Education proceeds from widely shared values, and we do, in fact, have widely shared values. We may not agree about how democracy is to be achieved and about whether we have too much or too little of it, but few would question the idea that each person has the right as a citizen to participate in the shaping of public issues. We believe in the idea of self-government and in the greatest possible involvement of citizens as voters, as volunteers in community organizations, as members of interest groups, and as spokesmen for different views. While we may differ over particular educational issues, there is general support for the idea that schooling is a necessary mechanism for achieving society's goals: to prepare the younger generation to be thoughtful citizens; to enable each person to appreciate and contribute to the culture; to sharpen the intellectual and aesthetic sensibilities for lifelong enjoyment; to develop readiness for the educational, occupational, and professional choices that each person will confront; to kindle a sense of responsibility for others and a sense of integrity; to teach children how to lead and how to follow; and to acquaint young people with the best models of achievement in every field while encouraging them to strive to realize their own potential.

If these are widely shared educational aims, and I believe they are, then none of them should be left to chance. The curriculum should be designed so that every student has the fullest opportunity to develop his powers, intelligence, interests, talent, and understanding. Every student needs to know how to form and formulate his own opinions. To do so, he must learn how to read critically, how to evaluate arguments, how to weigh evidence, and how to reach judgments on his own. Every student, to understand the world in which he will be a participant, should be knowledgeable about history; should master some other language as well as his own; should discover the pleasures of literature, especially its power to reach across time and cultures and to awaken our sense of universality; should study science and technology, both as a citizen who will be asked to comprehend complex issues and as an individual who must live with constant change. Since we believe that everyone should be equally concerned about the problems of our society, then we must believe that everyone, every student, should be schooled in a way that meets his need to know history, science, mathematics, language, the arts, literature, and so on. And yet it is not simply on the grounds of utility, relevance, and political value that the case for liberal education rests. We do not need to know how to read Shakespeare; we can be good citizens without any knowledge of Athenian civilization, even though our concept of citizenship is based on the very period of which we are ignorant. We must concern ourselves with the survival of history, philosophy, literature, and those other disciplines that may lack immediate utility because without them ours would be an intellectually impoverished and spiritually illiterate civilization.

To some people, all this is so self-evident that it ought not be necessary to plead for the value of an education of substance and content. Yet it is necessary, because of the widespread disarray in high school and college curricula. In the face of changes that have occurred in the past decade or so, many educators seem unable to remember how to justify or defend or champion liberal education. The proposition that all students should be subject to curricular requirements that define the essentials of a good education has become controversial, rather than a starting point in defining the nature of a good curriculum.

Confronted with conflicting demands from those who want reduced requirements and those who want curricular substance, many schools have resolved the dilemma by reducing requirements while expanding electives. Thus students may take history courses to meet their minimal graduation requirement, but may choose history courses that are little more than classes in current events. Or they may meet their English requirement by reading popular fiction, mystery stories, and science fiction. There is no harm in what is included; from the perspective of a liberal education, what is unfortunate is the wide body of knowledge that is excluded when course proliferation and lax requirements are joined together. Professors regularly encounter students who are ignorant of anything that happened before the Civil War as well as anything that happened, or was written, outside the United States. They may have heard of Plato and Aristotle in a survey course, but they have never read anything written by either and have only a dim notion (usually wrong) of what they "stood for." Mention Dickens, Tolstoy, Conrad, or Melville, and they have heard of them too, but they "didn't take that course." Some professors who teach literature have been astonished to find students who know nothing of mythology or the Bible; allusions to Job or Icarus must be explained to those who have no intellectual furniture in their minds, no stock of literary or historical knowledge on which to draw beyond their immediate experience. In a recent issue of *Commonweal*, J. M. Cameron soberly observes that if Freud attended school today, he might not be able to think up the Oedipus theory because he would not have enough mythology in his head to do so. We seem now to turn to television or the movies to teach the history and literature that were neglected in school. To permit knowledge to be fragmented, as we have, by serving it up cafeteria-style, with each person choosing whether to be minimally literate or to be a specialist, contributes to the diminution and degradation of the common culture.

AMERICAN COUNCIL ON EDUCATION
ONE DUPONT CIRCLE
WASHINGTON, D. C. 20036

OFFICE OF THE PRESIDENT

August 20, 1981

The Honorable Carl D. Perkins
Chairman, Committee on Education and Labor
U. S. House of Representatives
Washington, D. C. 20510

Dear Mr. Chairman:

On behalf of the American Council on Education, representing over 1,400 colleges, universities, and the higher education associations listed below we hereby request that this letter be included in the hearing record pertaining to HR 1662, The Educational Testing Act of 1981. As organizations whose members include nearly all of the colleges and universities in the nation, we are troubled by this proposed legislation and concerned about the impact its enactment would have on the admissions process at colleges and universities.

We recognize that faculty, students, parents, and others have a genuine and legitimate interest in the nature and use of the standardized tests that play a role in the admissions process at many higher education institutions. During the past four decades, such tests have proven to be useful at many institutions in assessing students' potential for completing undergraduate, graduate, and professional programs of study. In many instances, test results have been especially helpful to students who attended schools and colleges not well known to admissions committees and to those whose prior educational work did not reflect their full potential for further study.

At the outset we wish to make it clear that while acknowledging the interests of students and others in the nature and use of such tests and concurring with some of the stated goals of the proposers, we believe that Congress should not enact any legislation at this time. Portions of HR 1662 are unnecessary in that they simply codify the current practices of most testing programs and make no significant changes or improvements. Further, the proponents of HR 1662--which would require the disclosure of test questions and answers after each test administration--have not demonstrated the need for this major change in test program procedures on the basis of demonstrated educational benefits to students, on a definitive analysis of relative costs and benefits to students or institutions, or on well-documented instances of improper or harmful past practices. Finally, the portions of HR 1662 that require test agencies to make public all studies conducted on behalf of individual colleges would either impose unwarranted costs on colleges and universities or severely limit the ability of colleges to develop admissions procedures that best serve the needs of both students and institutions.

Historically, Congress has shown an admirable reluctance to legislate matters at the heart of education and curriculum. College admissions procedures (including the evaluation of test results) are an outgrowth of and a reflection of the mission and goals of each institution; the information required from each student applicant properly reflects the curriculum and requirements of the institution and the educational program to which application for admission is made. The subject of the proposed legislation is an educational matter and properly the province of the faculty of each institution. Although the proposers disclaim any intent to interfere with admissions procedures, it is our view that legislation which makes testing a more costly process or which may affect the usefulness of test results to colleges and universities does in fact interfere with admissions. Furthermore, the history of legislative and regulatory practices raises fears that this step threatens not only interference with one factor in the admissions process but makes all others vulnerable to further regulations.

The enactment of legislation in any area not previously subject to legislation suggests the existence of an undesirable and intolerable condition that can and must be remedied by law. Do such conditions exist in the area of educational admissions testing? We think not and do not believe the proponents of the bill now before you have carried their burden of demonstrating a clear need for such legislation.

Today, qualified students have greater access to higher education in the United States than ever before. More than 90 percent of the students currently seeking admission to undergraduate study are admitted to the college of their first choice; more than 95 percent are admitted to some college. As the number of 18-year olds declines, more and more undergraduate institutions can be expected to use test results for inclusive purposes to help identify and admit all students able to complete undergraduate programs of study rather than to exclude those who appear less likely to do so. At the graduate and professional levels, most institutions have already recognized the desirability of broadening their student bodies and have achieved diversity through vigorous programs of affirmative action, special academic assistance, and multiple criteria for admissions. This has occurred in the absence of legislation concerning the use of admissions tests.

HR 1662 prescribes specific items of information about each admissions test that must be provided to students when they register for the examination. This provision is unnecessary, in our judgment all the relevant information is already provided to students.

The sponsors of admissions tests regularly provide information about the nature of standardized tests and their content. The disclosure of test questions and their correct answers is now an operational policy of many major test sponsors. Those test agencies which have found it possible to do so have already instituted the disclosure of test materials throughout the nation. Test registration booklets, among other things, include information about the nature of these tests, the interpretation of test results, scoring procedures, standard error of measurement, validity, test disclosure, privacy of scores, test-score appeal procedures, and special services for handicapped students. In addition, reports and statistical data about testtakers, test revenues and expenses, where not confidential or precluded from release by a contracting agency, are already available to interested parties.

No persuasive evidence has been presented to suggest that test agencies are now handling student test scores improperly. To the best of our knowledge, test results are currently sent in identified form only to institutions and agencies designated by the student--either by a specific request or by a general permission applying to scholarship programs, special information services, and the like.

It is unwise to attempt to regulate the publication of research results. Results from studies of test validity that are of general interest and applicability should be and normally are published and widely distributed. Section 4 of HR 1662 seems an ineffective and objectionable way to accomplish this end. The requirement that researchers associated with admissions testing programs publish all of their work is at odds with commonsense and standards of good scholarship. Certainly, distinctions must be made between preliminary and final reports; between memoranda and more extensive treatments or analyses; between papers prepared to stimulate discussion and those that present conclusions resulting from the discussions.

Each year more than 1,000 studies of test validity and related matters are conducted by or on behalf of individual colleges and universities; each involves students enrolled at a single institution and each examines the usefulness of admissions test results in conjunction with other relevant information about the students as they relate to indicators of program completion and achievement at that institution. Currently, the results of these studies are available to faculty and admissions officers at that institution for their use in developing appropriate admissions procedures. None is of interest beyond that institution; none necessarily has relevance for students at other institutions. Yet, section 4 of HR 1662 requires that all studies of test validity be made public if they have been conducted by a test agency on behalf of the institution or if they have used data obtained from a test agency.

Individual institutions should be encouraged to conduct such studies. The public filing requirement of section 4 of HR 1662, however, is likely to discourage these efforts. The test agency studies prepared for an institution consist largely of tabular material. They lack the interpretive context that is known to admissions committees and therefore would be either unintelligible or uninformative to other readers. It would serve no purpose to require the publication of such studies.

In developing test questions the threats of ambiguity, bias, and faulty reasoning are always present--whether the writer is a teacher, scholar, or professional item writer. All test agencies conduct a review of admissions test items prior to their first use with students in order to minimize the possibility that these threats to test validity will obscure the actual abilities of the students. The review conducted by persons familiar with the test content and skills is an important step in the development of quality examinations, but many persons are not familiar with the procedures used or with the qualifications of the reviewers. Perceptions that admissions testing programs are shrouded in mystery and secrecy could be changed if test sponsors prepared and distributed detailed descriptions of the review process used in developing their tests. The prompt move to correct errors on national exams in recent months has gone a long way in further instilling confidence in the equity of the testing system. Test agencies might give consideration to the more widespread adoption of the practice of involving qualified reviewers nominated by national associations of teachers and faculty.

We appreciate the opportunity to present our views to your committee and stand ready to respond to any questions that you or your staff may have.

Cordially,

J. W. Peltason

J. W. Peltason

The following associations wish to join in this statement:

American Association of State Colleges and Universities
American Council on Education
Association of American Colleges
Association of American Universities
Association of Catholic Colleges and Universities
Association of Jesuit Colleges and Universities
Council of Graduate Schools in the United States
National Association of Independent Colleges and Universities
National Association of State Universities and Land-Grant Colleges

cc: Members of the Committee

AMERICAN MEDICAL ASSOCIATION
535 NORTH DEARBORN STREET • CHICAGO, ILLINOIS 60610 • PHONE (312) 751 6000 • TWX 910 221 0300

JAMES H. SAMMONS, M.D.
Executive Vice President
(751 6200)

August 4, 1981

The Honorable Carl D. Perkins
Chairman,
Committee on Education and Labor
2181 Rayburn House Office Building
Washington, D.C. 20515

RE: H.R. 1662,
"The Educational Testing Act of 1981"

Dear Mr. Chairman:

The American Medical Association takes this opportunity to register its opposition to H.R. 1662. This proposed legislation mandates that an individual who had taken a standardized test that was administered to 5,000 or more people would be able to receive a copy of the test, the correct answers, and his or her answer sheet. In addition, information concerning the purpose and intended use of the test, the subject matter of the examination and what it purports to measure, and the reliability of the exam would be provided to potential test takers.

The AMA supports the elements of the bill that would help establish requirements to explain the purpose and use of standardized testing, and that would allow the Secretary of Education statistically to analyze the results of such testing. The Association is concerned, however, that H.R. 1662 would adversely affect the quality of standardized tests, and in particular, the quality of the Medical College Admissions Test (MCAT) to the detriment of schools, students, and the public.

One of the major problems of releasing test questions is the necessity of developing new questions for subsequent testing. In developing standardized tests that are designed to measure narrowly defined areas of knowledge, such as the scientific segment of the MCAT, the potential for the creation of valid questions is not nearly as extensive as exists in developing a test to measure general knowledge or verbal ability. Furthermore, to create a statistically valid question for a standardized test, a common procedure is to pretest the question within an examination without considering the answer in determining the test taker's score.

The disclosure provisions of H.R. 1662 would make it impossible to maintain the integrity and validity of test questions, including the pretest questions. New questions would have to be developed without the customary pretesting, and this could adversely impact upon the validity of the examination as a testing device.

While the AMA is not involved in test development or administration, we have long sought to encourage the highest standards of medical education. Those standards can best be met by students who are qualified to undergo a long and rigorous course of medical education. Even with the variety of measuring criteria, it is difficult to evaluate a student's potential for academic success in a medical education program. Admission to medical school is determined by many factors, and not based solely on the basis of a standardized examination. Other factors used in evaluating potential medical school students include undergraduate academic performance, personal interviews and faculty recommendations. Nonetheless, testing is very important in the admissions process because it is the single measure that is independent of local variables, such as degree of undergraduate program difficulty, institutional grading practices and standards, and the academic ability of the applicant's peer group.

Coupled with individual judgment, experience, and creativity, a physician must utilize scientific knowledge, observational skills, and problem-solving ability to provide medical care properly. To the degree that standardized testing can aid in identifying any of these characteristics, the tests do serve a valid purpose. The imposition of specific disclosure requirements could make it extremely difficult to develop and administer tests that will be effective in achieving their objectives. We are concerned that the disclosure requirements could prove detrimental to an established methodology that has been successfully utilized in the process of selecting qualified medical students.

We urge the Subcommittee not to adopt H.R. 1662.

Sincerely,

James H. Sammons, M.D.

JHS:jg
0081p

THE ASSOCIATION OF AMERICAN VETERINARY MEDICAL COLLEGES

July 24, 1981

The Honorable Carl D. Perkins, Chairman
Subcommittee on Elementary, Secondary,
  and Vocational Education
Committee on Education and Labor
U.S. House of Representatives
B-346-C Rayburn House Office Bldg.
Washington, D.C. 20515

Dear Mr. Perkins:

Enclosed is a statement of the Association of American Veterinary Medical Colleges on H.R. 1662. We regret that we could not request an opportunity to present our views at the hearings of your subcommittee and the Subcommittee on Postsecondary Education on July 21 and 22.

Since we are vitally concerned with this matter, we would appreciate having our statement included in the record of the hearings. We will also provide it to the members of both subcommittees. Thank you.

Sincerely,

E O Dickinson

Earl O. Dickinson, D.V.M.
President

Enclosure

1522 K STREET, N.W. — SUITE 828 — WASHINGTON, D.C. 20005 — 202/659-2040

STATEMENT
OF THE
ASSOCIATION OF AMERICAN VETERINARY MEDICAL COLLEGES
FOR THE RECORD OF HEARINGS
OF THE
SUBCOMMITTEES
ON
POSTSECONDARY EDUCATION
AND
ELEMENTARY, SECONDARY AND VOCATIONAL EDUCATION
OF THE
COMMITTEE ON EDUCATION AND LABOR

HOUSE OF REPRESENTATIVES

JULY 21-22, 1981

The Association of American Veterinary Medical Colleges appreciates this opportunity to present its views on H.R. 1662, the "Educational Testing Act of 1981."

"Standardized" examinations are given for a number of reasons. A few of the very important ones are: (1) to establish "norms" and identify those who, for whatever reason, are at the extremes; (2) to provide some long-term and continuing basis for comparisons which may be useful in evaluating the total system; (3) to identify "areas" of strength and weakness in the examinees both individually and collectively; and (4) to provide another element in the process used for selection. Admissions groups need all the help they can get in making sound judgments on selections to highly competitive programs, and licensing boards must determine competency to perform and render reliable services to the consuming public. In a purely academic setting, the educational value of providing feedback on examinations has proponents and opponents, but in that setting there is general recognition that such procedures do have educational value. However "Standardized Testing" as currently conducted, whether it be for entry to the university or for professional licensure decisions, has different objectives. The state and national dimensions of these examinations also imply an entirely different set of circumstances than that for an examination in a given course in a given educational institution, and detailed feedback is not essential.

Specifically, the Association of American Veterinary Medical Colleges speaks against the proposed legislation because:

(1) We are not convinced that such proposals would solve any ills that are perceived as currently existing. In fact, such legislation would be complicating and counterproductive to the overall objectives of standardized testing as currently practiced.

(2) The body of knowledge is extensive and sampling is essential. As major and important concepts are tested for in examinations, new test items would tend toward the trivial if questions and answers were routinely published.

(3) Flexibility would be greatly reduced. Students now have an opportunity to take a number of examinations in varied locations at various times. For example, the Veterinary Aptitude Test (VAT) is currently offered 5 times per year. There would be a move to many fewer offerings per year and in the instance of VAT probably to once per year.

(4) The impact would be inversely proportional to the size of the group affected. For example, most high school students in the United States take the SAT and/or ACT test, and there is a relatively low development cost per student as compared with those tests for many, many fewer examinees, e.g., human medicine and MCAT and veterinary medicine and VAT where the development cost would increase significantly.

(5) There would be a great waste of professional time in the redundancy of new examination construction, detracting from the ability to evaluate, analyze, and improve.

(6) The tendency to produce a "levelling effect" in test performance is a grave concern, since the advantages of innate ability and concerted study effort on a broad front would tend to be negated.

Fairness, objectivity, and capriciousness are not in contention as these are fundamental and obvious concerns of society. Examinees need to know the nature and use of examinations and to be assured that test scores accurately reflect performance. Feedback on areas of strength and weakness should be available. We believe testing services are sensitive to these matters and are not just the money-mongers some try to make them out to be. However, the profit motive is as real in providing services as it is in selling hard merchandise, and the free enterprise system, if it is to obtain, cannot be battered and beat by the very hand that it feeds.

Evaluation is as fundamental to the educational system as the jury is to the judicial system. An examination, which is an evaluation device, is an exercise to determine a person's knowledge, ability, and qualifications in a prescribed area which may be rather broad or rather limited in scope. If a person has both the questions or problems and the answers, there is in effect no examination, and the fundamental objective of the whole process—evaluation—is negated. That the system is imperfect is recognized and admitted, but so is the political system, the judicial system, or any other system devised by imperfect people. Before the system is radically manipulated it should be understood as it is being used and with the full recognition of what radical changes will accomplish.

As participants and educators in a health profession field, we support the concept of standardized testing. Concurrently it should be emphasized that such tests are not the only criterion used in arriving at judgment decisions which must be made. Rather, test results are combined with other academic and sociological data to provide a breadth of input in helping to evaluate ability and motivation. Discriminating elements need to be honed and refined and not blunted or made more diffuse.

The Association of American Veterinary Medical Colleges can see no benefit to students, the general public, educational institutions, or the educational testing "industry" from this bill, and we question whether the bill would have the effects claimed by its proponents. We believe it is necessary to reject this bill in order to preserve the systems of graduate professional admissions as they now exist.

THE EDUCATIONAL TESTING ACT OF 1981

WEDNESDAY, NOVEMBER 4, 1981

HOUSE OF REPRESENTATIVES, SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION, JOINTLY WITH THE SUBCOMMITTEE ON POSTSECONDARY EDUCATION, COMMITTEE ON EDUCATION AND LABOR,
Washington, D.C.

The subcommittee met, pursuant to adjournment, at 10:00 a.m., in room 2175, Rayburn House Office Building, Hon. Ted Weiss presiding.

Members present: Representatives Weiss, Erdahl, Craig, Williams, and Kildee.

Staff present: John F. Jennings, counsel; William Blakey, counsel; and Nancy L. Kober, legislative specialist.

Mr. WEISS. We will commence and if everyone will take their seats we will proceed.

Let me at the outset indicate that as usual there are more things happening around here than there are members to do them. As a result with King Hussein and the Budget Committee hearings and so on, our other members will be drifting in. Indeed, the chairman of the Postsecondary Education Subcommittee, Mr. Simon, has to be at a suddenly called meeting of the Budget Committee and he will probably not be able to be here this morning but asked that his opening statement be entered into the record and that will be done at this stage of the proceeding without objection.

[Opening statement of Chairman Simon follows:]

OPENING STATEMENT OF HON. PAUL SIMON, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF ILLINOIS AND CHAIRMAN OF THE SUBCOMMITTEE ON POSTSECONDARY EDUCATION

This past summer, the Subcommittee on Elementary, Secondary, and Vocational Education and the Subcommittee on Postsecondary Education held 2 days of joint hearings on the Educational Testing Act of 1981 (H.R. 1662) introduced by Congressman Ted Weiss of New York. Interest by this body in the issue of Truth in Testing began with the introduction of the Truth in Testing Act of 1979 (H.R. 3564) by Mr. Gibbons, our colleague from the State of Florida. Many of us have participated in a number of hearings throughout the years so that we could become fully aware of the issues, and the pros and cons related with the enactment of test disclosure legislation at the national level.

The joint hearings that we held this past July, focused on testimony regarding the field of test and measurement; how tests are utilized; changes occurring in the area of testing; and, experience test takers of standardized tests have had. We have heard the amount of time entailed in the development of tests, the manner in which institutions of higher education utilize standardized tests for admissions purposes, and what test scores on standardized tests like the SAT mean to students and parents applying to college for the first time.

Interest continues in the issues of test disclosure and the nature and quality of standardized tests among legislators, the public, the testing industry, and those in the education field. In September, the California Legislature enacted a test disclosure bill. The Law School Admission Council has announced that beginning in 1982 the LSAT will be modified to deemphasize mathematical skills and provide a writing sample. An articile in the New York Times in which this announcement appeared and a later article in the Chronicle of Higher Education also touched on the steps being taken to examine the utilization of alternative modes of assessment for admissions to institutions of higher education.

In an attempt to gather further data and allow for public input as to the direction this body should take in regards to the Truth in Testing issue, we have invited Gregory Anrig, the president of the Educational Testing Service and Dr. Philip Rever, the director of the American College Testing Service to provide testimony regarding their policies with regard to disclosure and provide their views on the pending legislation. We also have here present today, Ms. MariAnn Austin, representing State Senator Kenneth LaValle of New York to present testimony on the test disclosure legislation enacted in New York. In addition, John A.D. Cooper, president of the Association of Medical Colleges is present to discuss his organization's experience with test disclosure and their filing for an injunction prohibiting the enforcement of the New York statute against the MCAT.

The testimony that these experts in the field of testing and legislation will provide us today will assist us in our deliberations regarding H.R. 1662.

Mr. WEISS. As sponsor of the legislation which is our topic today, I again want to express my appreciation to both subcommittees for conducting these hearings. Previous hearings on the Educational Testing Act of 1981 have been instrumental, I think, in shaping the debate on the issue of open testing as well as precipitating much of the activity which has occurred in the field of standardized testing since this legislation was introduced in 1979. I am certain that the testimony we receive today and tomorrow will show the continued importance of public discussion of accountability and fairness in standardized testing.

Since congressional hearings began on this topic in July 1979, indeed since the adoption of legislation in the State of New York that same year, nothing short of a minor revolution has occurred. Major testing agencies have revamped their procedures for disclosing some information on standardized exams and this has heightened public concern about these test groups. Perhaps more importantly a fundamental shift seems to be underway in the attitude test takers everywhere bring to their standardized admission test. No longer are test takers routinely blocked from obtaining information which might be both helpful and educational. These tests now are finally being made subject to the public scrutiny they deserve.

The real measure of the distance traveled can perhaps be seen by the fact that in 1957 students were not permitted to even know their test scores, which were reported blindly to colleges and universities. And until 1980 students were unable to see their own test sheets. Now at least one major testing agency has decided that all test takers should receive their test sheets, a listing of correct answers and the test questions as a matter of course. This testing agency, the Law School Admissions Council, and others which have implemented limited disclosure policies should be commended. Yet in hearings held earlier this year there was widespread agreement that most of the changes made by testing agencies since July 1979, would not have occurred had it not been for legislative and congressional involvement.

I continue to welcome suggestions and observations on H.R. 1662. The experience with open testing statutes in New York State and elsewhere has shown that some refinements may well be appropriate in that bill.

Let me conclude by noting that the issue we are addressing in this round of hearing is fundamentally the same issue we focused on when the hearings began in 1979, insuring that testing agencies are held appropriately accountable to the millions of students they serve each year.

We have four important witnesses scheduled for this morning. They will be questioned as members of the panel deem it appropriate and then we will have a second set of hearings starting tomorrow morning at 9:30.

The first witness who we're pleased to welcome is the new president of the Educational Testing Service, Mr. Gregory Anrig.

STATEMENT OF GREGORY ANRIG, PRESIDENT, EDUCATIONAL TESTING SERVICES

Mr. Anrig. Mr. Chairman and members of the subcommittee, I am deeply grateful for the privilege of appearing before your committee today. I would have to confess that as I rounded the corner and came down the hall today and saw the guards and the TV lights and crowds of people out there I said, "My God, they didn't tell me it was going to be like this." I was pleased to find out that I was deferring that privilege to the King of Jordan and hope that our hearings will be less eventful than that one obviously is going to be.

My name for the record is Gregory R. Anrig, and I am president of the Educational Testing Service. I appreciate very much, Mr. Chairman, your opening comments in which you recognize the responsiveness of test organizations over the course of the last couple of years. I too believe they have been responsive. I also believe and wish to recognize that these hearings have served a worthy purpose in prompting and guiding that responsiveness; and, Mr. Chairman, I recognize that you yourself had a very important roll in that and would like to be sure to acknowledge that.

I have only held the position of president of the Educational Testing Service since September 1 of this year. That automatically means I'm an expert in testing by definition, but I would have to confess that only for two months have I been so. In appointing me president, the ETS board of trustees selected a person whose entire professional life has been in public service at the local, State, and Federal levels of government. My two predecessors, as you know, the first came out of higher education and was in effect a builder, Henry Chauncey. My second predecessor, my immediate past predecessor, was an expert and a scholar in testing and has grown up really through the testing organizations both college board and ETS. In contrast to that, for the last 8½ years I've served as commissioner of education for the Commonwealth of Massachusetts.

My selection as ETS president reflects the judgment of the board of trustees about where Educational Testing Service is as a private, not-for-profit educational organization and where it will be heading. This judgment also reflects, to some degree, the trustees' view of the future of educational testing in America generally.

I am committed to public service, the kind of public accountability you called for, and the fundamental principle of equal educational opportunity. My actions in each of these areas, because I've been a public official, are a matter of public record. A record on which I stand proudly, and can be attested to by any member of the distinguished Massachusetts delegation to the House of Representatives.

I don't shy from public accountability. I welcome it. It has been and is my position that anything has the potential to affect the life chances of youths and adults should be subject to reasonable public accountability. I believe that includes educational testing.

As education commissioner in Massachusetts, I testified in 1980 before the Massachusetts Legislature, it may surprise you, Mr. Chairman, in favor of a so-called truth-in-testing bill on behalf of the State board of education. But I want to point out that that testimony was conditional. It was my position then and that of the State Board of Education that public accountability in educational testing was best achieved by voluntary action of testing organizations. We urged them to take the kinds of actions which subsequently they have taken. As one with long experience in the development and carrying out of law, I feel that a democratic society should legislate only that which is absolutely necessary.

The question before your distinguished subcommittees of Congress and before oganizations which conduct educational testing is how best to achieve public accountability in educational testing in general, and specifically whether greater government intervention is necessary to accomplish this end.

I felt it was important that I used the privilege of appearing before this joint hearing today as an opportunity to present an answer to that question. To assure you that my testimony represents the corporate position and committment of Educational Testing Service, I recommended and on October 6, 1981 the ETS board of trustees unanimously approved a trustee policy on public issues which I am proud with your permission to submit for the record today as the basis for my testimony which follows.

What you will find newest and what I will describe is the willingness of the ETS board of trustees and the new ETS president to raise these actions to the level of corporate policy commitment, to directly involve the board of trustees itself as the final accountable body for ETS, and to invite outside examination both of our operations and of our data.

This policy grew out of intensive discussions among the trustees and among the senior staff of ETS. Although I consulted with representatives of each of our client boards in developing the recommended policy, I wish to make clear that my testimony today represents only the corporate position of Educational Testing Service. Other organizations, including our client boards, have presented and will present their views independent of the position I take today.

The issue of accountability before us, in my judgment, is not "truth" in testing. My review of the many arguments advanced in the debates of the last several years demonstrates very clearly to me that no sides in these debates had a patent on truth. The real issue is openness in testing as the chairman has expressed in the opening statement.

I believe it is possible to achieve openness in testing voluntarily, without Government intervention, and ETS commits itself firmly to that principle.

Openness in testing requires taking a position in the public's interest, having standards for this position by which you are willing to be judged, having a way to enforce those standards, and creating a means to review and report publicly upon these actions in a manner that has credibility.

As my first action after taking office on September 1, I recommended that we voluntarily take such actions to demonstrate our commitment and to do so not by presidential decision but by formal policy of our governing board.

On October 6 the ETS board of trustees formalized its support for the principle of openness in admissions testing by official vote. Specifically, the trustees endorse the recent actions taken on release of tests and test answers by the college board, the graduate management admission council, the graduate record examinations board, the law school admission council and the TOEFL policy council, that is the test of English as a foreign language. Further, by vote of the ETS trustees, those actions and commitments are incorporated as policy for ETS in its services for the respective boards and will govern ETS in the administration of those test programs.

Openness in testing, however, and I think this is important for us to be clear on, involves far more than just test disclosure. ETS, therefore, also has adopted as policy the public interest principles proposed for discussion 2 years ago by ETS and the major admissions testing boards with which we work, and I submit a copy today for the record if the Chair approves.

These public interest principles affirm our commitment to the rights of test takers in admission testing programs: The right to know what is in the test, to have fair and bias-free test, to receive accurate test scores, to have access to tests regardless of handicap or other special circumstances, to be informed about what the tests and their scores mean, to be protected in the privacy of information about themselves, to the promotion of proper test score use.

In adopting the public interest principles, I believe we are the first national testing organization to raise these principles from the level of discussion to the level of corporate policy commitment. I know and you know that the ETS position which I have just stated requires specific standards if that general commitment is to have real meaning. We, therefore, have developed a very comprehensive set of principles and policies to govern ETS tests and services. They have been tested as guidelines which are now by trustee vote in October of this year to be called the ETS standards for quality and fairness, which I also submit for the record. The ETS standards for quality and fairness are designed to insure that our products and services measure up to clear and strong specifications in the areas of: accountability, protection of privacy, product accuracy, test quality, responsible use of test results, and research and technical assistance. These standards are fully consistent with and in many respects exceed the current joint testing and measurement standards of the American Psychological Association, the American Educational Research Association and the National Council on Measurement in Education. As you know, Mr. Chairman, those are the three groups which are recognized nationally for standard setting in the area of tests and measurement.

ETS will hold itself accountable to these demanding standards and the board of trustees has committed our organization formally to do so.

With regard to enforcement of these standards, ETS has developed and implemented a comprehensive audit process to assure that all testing programs adhere to the ETS standards for quality and fairness. These audits will review each ETS-developed testing program at least once every 3 years, a process already begun. The audits will be conducted under specific oversight by the president and general oversight by the ETS trustee committee on public responsibility, and I'm already carrying out my responsibilities in that area. If a testing program does not meet these standards, I pledge ETS to work diligently with the program sponsor to bring the test into compliance with the standards.

The trustee committee on public responsibility has been charged by the ETS board of trustees to prepare and publicly issue on an annual basis an accountability report on actions taken to adhere to the ETS standards for quality and fairness. To aid the trustee committee on public responsibility, a visiting committee will be formed by persons from outside of ETS and its board of trustees who are knowledgeable in the areas addressed by the ETS standards for quality and fairness. It is my full intent to include on the visiting committee persons from organizations which have in the past been critical of ETS. That, by the way, is old practice for this particular person. I've always felt it was better to invite them in than to hold them out.

A series of actions will be taken to broaden access to ETS itself. While complaints from students and other test takers currently are responded to earnestly at ETS and I've been very impressed by that, the process for responding differs from one testing program to another. A uniform complaint procedure will be implemented across all ETS-developed testing programs, closely coordinated with each of our client boards to improve the way we can respond to student complaints. I also have just designated an ETS student concerns coordinator in the office of the president to monitor this procedure and to provide communications with students. She is a highly competent professional staff person who already has earned my confidence and whose office is next to mine for easy access, and she also I might add participates in all of the major management planning processes of the organization. It will be her job to assure that student concerns both are advocated and brought to those areas of deliberation.

Actions have been initiated to increase further the variety of information about ETS tests available to individuals, schools, community agencies and institutions of higher education. These will include new formats, such as audiovisual materials, videotapes for public and cable television, and handbooks aimed at accurately informing test takers and the public about standardized tests, the skills which these tests measure, and what one should know about taking the tests. The aim of these materials is to make readily available to all students, regardless of economic status, high quality information about the tests developed by ETS.

Finally, the trustee committee on research and development has been working as late as yesterday afternoon to broaden and simplify access to ETS-controlled test and research data, consistent with our longstanding commitment to protect the privacy of individual test takers. Recommendation to make these resources more accessible to researchers and others through public use tapes will be presented by the committee to the board of trustees in April 1981 and I anticipate speedy action on this important matter. We intend, also, to work collaboratively with our contracting boards to facilitate easier access to the data which they control in accordance with their policies in this area.

The five elements which I have described and which are embodied by corporate vote in the ETS trustee policy on public issues, demonstrate our commitment to openness in testing. These elements—an affirmative position on such openness, standards of accountability, enforcement of these standards, public reporting and organizational accessibility—can and will be broadened and strengthed over time, in close collaboration with the organizations which contract for ETS services. We commit ourselves to that goal.

I am confident that such voluntary action is better than Federal legislation, Mr. Chairman. I don't have to describe to you and the members of your committee the reams of regulations, forms, and bureaucratic guidelines which in the past have been sired by legislation such as you now have before you regarding educational testing. Legislation of this type inevitably leads to greater governmental intrusion into education, and will needlessly increase the costs of educational testing for thousands of test takers and their parents who are your constituents.

The goal of openness in testing can be and is being achieved without further legislation. In my judgment that is how it should be. The New York State law is being carried out diligently by ETS and its test sponsors, and a summary of these actions is submitted for the record with the Chair's permission. By voluntary decision of the respective test taking boards, the basic provisions of this law have been applied on a national basis. Further legislation such as H.R. 1662 is not needed or warranted.

To those who fear retrenchment from voluntary openness in testing, I say judge ETS and others both by our words and our actions now and in the future and act accordingly. Don't legislate for the future to remedy problems of the past which largely have been addressed. Don't impose regulation on the entire industry when the major segments of that enterprise have already acted voluntarily to meet the major goals of that legislation. If there is retrenchment then you can act. If there isn't, why act at all?

ETS is firmly and fully committed to the principle of openness in testing. To advance this principle for the future, I offer a challenge and a proposal. I urge all profit and nonprofit testing corporations, and those agencies of Government which test, to join with us voluntarily in an industrywide "Code of Fair Testing" for educational testing nationwide. I recommend that this voluntary Code of Fair Testing be based on the new standards now being developed by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. I further propose that a panel be appointed jointly by these three associations to monitor the industrywide Code of Fair Testing and I commit ETS today to cooperate fully with such a panel.

Thus, in addition to the actions which I've described affecting educational testing service, the educational testing service board of trustees and I offer to join other testing corporations and agencies in a national effort to voluntarily promote openness in testing.

In the 2 months I've been at ETS I've been very impressed by the people who make up the organization and by the quality of services they provide to individuals and to institutions. ETS always has stood for quality and equality of educational opportunity and if it didn't I wouldn't have gone there. It is dedicated to continuing this tradition in the challenging years ahead. I hope my testimony today and the policy position of the ETS board of trustees make this commitment and this dedication clear.

I thank you, Mr. Chairman, and your colleagues on the subcommittee for the privilege of appearing before you today. I shall be glad to respond to your questions.

Mr. WEISS. Thank you very much, Commissioner. Let me at the outset commend you as the president of the educational testing service and tell you how welcome it is to have the attitude which is implicit throughout your testimony. I think that underlying any possibility of success in this or any other endeavor that we undertake legislatively is the attitude of the people who are involved, and I think indeed there is a new wind blowing and I welcome it.

I also want, without objection, to have all the appendices and extraneous material that you referred to entered into the record so that we'll have all that for our perusal and for any others who want to look at the record of this hearing.

[The information referred to follows:]

PREPARED STATEMENT OF GREGORY R. ANRIG, PRESIDENT, EDUCATIONAL TESTING SERVICE

Mr. Chairman:

My name is Gregory R. Anrig and I am President of Educational Testing Service (ETS).

I have held this position only since September 1 of this year. In appointing me President, the ETS Board of Trustees selected a person whose entire professional life has been in public service at the local, state, and federal levels of government. For the last 8 1/2 years, I have served as Commissioner of Education for the Commonwealth of Massachusetts.

My selection as ETS President reflects the judgment of the Board of Trustees about where Educational Testing Service is, as a private, not-for-profit educational organization, and where it will be heading. This judgment also reflects, to some degree, the Trustees' view of the future of educational testing in America generally.

I am committed to public service, public accountability, and the fundamental principle of equal educational opportunity. My actions in each of these areas are a matter of public record--a record on which I stand proudly--and can be attested to by any member of the distinguished Massachusetts delegation to the House of Representatives.

I do not shy from public accountability. I welcome it. It has been and is my position that anything that has the potential to affect the life chances of youths and adults should be subject to reasonable public accountability. I believe that includes educational testing.

As Education Commissioner in Massachusetts, I testified in 1980 before the Massachusetts Legislature in favor of a so-called "Truth-in-Testing" bill, on behalf of the State Board of Education. But that testimony was conditional. It was my position and that of the State Board of Education that public accountability in educational testing was best achieved by voluntary action of testing organizations. We urged them to take the kinds of actions which subsequently they have taken. As one with long experience in the development and carrying out of law, I feel that a democratic society should legislate only that which is absolutely necessary.

The question before your distinguished Subcommittees of Congress and before organizations which conduct educational testing is how best to achieve public accountability in educational testing in general, and specifically whether greater government intervention is necessary to accomplish this end.

I felt it was important that I use the privilege of appearing before this joint hearing today as an opportunity to present an answer to that question. To assure the Subcommittees that my testimony represents the corporate position and commitment of Educational Testing Service, I recommended and on October 6, 1981 the ETS Board of Trustees unanimously approved a Trustee Policy on Public Issues (Appendix A) which I am proud to submit for the record today as the basis for my testimony which follows.

This policy grew out of intensive discussions among the Trustees and among the senior staff of ETS. Although I consulted with representatives of each of our client boards in developing the recommended policy, I wish to make clear that my testimony today represents only the corporate position of Educational Testing Service. Other organizations, including our client boards, have presented and will present their views independent of the position I take today.

The issue of accountability before us, in my judgment, is not "truth" in testing. My review of the many arguments advanced in the debates of the last several years demonstrates to me that no sides in these debates had a patent on truth. The real issue is openness in testing.

I believe it is possible to achieve openness in testing voluntarily, without government intervention, and ETS commits itself firmly to that principle.

Openness in testing requires taking a position in the public interest, having standards for this position by which you are willing to be judged, having a way to enforce those standards, and creating a means to review and report publicly upon these actions in a manner that has credibility.

The ETS position. On October 6, 1981 the ETS Board of Trustees formalized its support for the principle of openness in admissions testing by official vote. Specifically, the Trustees endorse the recent actions taken on release of tests and test answers by the College Board, the

Graduate Management Admission Council, the Graduate Record Examinations Board, the Law School Admission Council and the TOEFL* Policy Council. Further, by vote of the ETS Trustees, those actions and commitments are incorporated as policy for ETS in its services for the respective boards and will govern ETS in administration of those test programs.

Openness in testing, however, involves more than just test disclosure. ETS therefore also has adopted as policy the "Public Interest Principles" (Appendix B) proposed for discussion two years ago by ETS and the major admissions testing boards with which we work.

These "Public Interest Principles" affirm our commitment to the rights of test-takers in admissions testing programs:

-- to know what is in the test

-- to have fair and bias-free tests

-- to receive accurate test scores

-- to have access to tests regardless of handicap or other special circumstances

-- to be informed about what the tests and their scores mean

-- to be protected in the privacy of information about themselves

-- to the promotion of proper test score use

In adopting the "Public Interest Principles," I believe we are the first national testing organization to raise these principles from the level of discussion to the level of a corporate policy commitment.

---

*Test of English as a Foreign Language

Standards for accountability. The ETS position which I have just stated requires specific standards if that general commitment is to have real meaning. We therefore have developed a very comprehensive set of principles and policies to govern ETS tests and services. They have been tested as guidelines which are now, by Trustee vote in October of this year, the "ETS Standards for Quality and Fairness" (Appendix C).

The "ETS Standards for Quality and Fairness" are designed to ensure that our products and services measure up to clear and strong specifications in the areas of: accountability, protection of privacy, product accuracy, test quality, responsible use of test results, and research and technical assistance. These Standards are fully consistent with and, in many respects, exceed the current joint testing and measurement standards of the American Psychological Association, the American Educational Research Association and the National Council on Measurement in Education. ETS will hold itself accountable to these demanding standards and the Board of Trustees has committed our organization formally to do so.

Enforcement of standards. ETS has developed and implemented a comprehensive audit process to assure that all testing programs adhere to the "ETS Standards for Quality and Fairness." These audits will review each ETS-developed testing program at least once every three years, a process already begun. The audits will be conducted under specific review by the President and general review by the ETS Trustee Committee on Public Responsibility. If a testing program does not meet these standards, I pledge ETS to work diligently with the program sponsor to bring the test into compliance with the standards.

Public reporting. The Trustee Committee on Public Responsibility has been charged by the ETS Board of Trustees to prepare and publicly issue annually an accountability report on actions taken to adhere to the "ETS Standards for Quality and Fairness." To aid the Trustee Committee on Public Responsibility, a Visiting Committee will be formed of persons from outside of ETS and its Board of Trustees who are knowledgeable in the areas addressed by the "ETS Standards for Quality and Fairness." It is my intent to include on the Visiting Committee persons from organizations which have in the past been critical of ETS.

Openness of the organization and its information. A series of actions will be taken to broaden access to ETS itself. While complaints from students and other test-takers currently are responded to earnestly at ETS, the process for responding differs from one testing program to another. A uniform complaint procedure will be implemented across all ETS-developed testing programs, closely coordinated with each of our client boards, to improve the way we can respond to student complaints. I also have just designated an ETS Student Concerns Coordinator in the Office of the President to monitor this procedure and to promote communication with students.

Actions have been initiated to increase further the variety of information about ETS tests available to individuals, schools, community agencies and institutions of higher education. These will include new formats such as audiovisual materials, videotapes for public and cable television, and handbooks aimed at accurately informing test-takers and the public about standardized tests, the skills which these tests measure, and what one should know about taking the tests. The aim of these materials is to make readily available to all students, regardless of economic status, high quality information about tests developed by ETS.

Finally, the Trustee Committee on Research and Development has been working to broaden and simplify access to ETS-controlled test and research data, consistent with our long-standing commitment to protect the privacy of individual test-takers. Recommendations to make these resources more accessible to researchers and others will be presented by the Committee to the Board of Trustees in April 1982 and I anticipate speedy action on this important matter. We intend, also, to work collaboratively with our contracting boards to facilitate easier access to the data which they control, in accordance with their policies in this area.

The five elements which I have described, and which are embodied by corporate vote in the ETS Trustee Policy on Public Issues, demonstrate our commitment to openness in testing. These elements--an affirmative position on such openness, standards of accountability, enforcement of these standards, public reporting and organizational accessibility--can and will be broadened and strengthened over time, in close collaboration with the organizations which contract for ETS services. We commit ourselves to this goal.

I am confident that such voluntary action is better than federal legislation. I don't have to describe to your Subcommittees the reams of regulations, forms, and bureaucratic guidelines which, in the past, have been sired by legislation such as you now have before you regarding educational testing. Legislation of this type inevitably leads to greater governmental intrusion into education, and will needlessly increase the costs of educational testing for thousands of test-takers and their parents who are your constituents.

The goal of openness in testing can be and is being achieved without further legislation. That is how it should be. The New York State law is being carried out diligently by ETS and its test sponsors (see Appendix D). By voluntary decision of the respective test-governing boards, the basic provisions of this law have been applied on a national basis. Further legislation such as H.R. 1662 is not needed or warranted.

To those who fear retrenchment from voluntary openness in testing, I say judge ETS and others both by our words and actions, now and in the future, and act accordingly. Don't legislate for the future to remedy problems of the past which largely have been addressed. Don't impose regulation on an entire industry when the major segments of that enterprise already have acted voluntarily to meet the major goals of that legislation. If there is retrenchment, then act. If there isn't, why act at all?

A Challenge and a proposal. ETS is firmly and formally committed to the principle of openness in testing. To advance this principle for the future, I offer a challenge and a proposal:

— I urge all profit and nonprofit testing corporations, and those agencies of government which test, to join with us voluntarily in an industry-wide "Code of Fair Testing" for educational testing nationwide.

-- I recommend that this voluntary "Code of Fair Testing" be based on the new standards now being developed by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education.

-- I further propose that a panel be appointed jointly by these three associations to monitor the industry-wide "Code of Fair Testing" and I commit ETS today to cooperate fully with such a panel.

Thus, in addition to the actions which I have described affecting Educational Testing Service, the ETS Board of Trustees and I offer to join other testing corporations and agencies in a national effort to voluntarily promote openness in testing.

I am new to ETS. I have been very impressed by the commitment of the people who make up the organization and by the quality of services they provide to individuals and to institutions. ETS always has stood for quality and equality of educational opportunity and is dedicated to continuing this tradition in the challenging years ahead. I hope my testimony today and the policy position of the ETS Board of Trustees make this dedication clear.

I thank you, Mr. Chairman and, through you, your colleagues on the Subcommittees, for the privilege of appearing before you today. I shall be glad to respond to your questions.

Appendix A

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08541

Board of Trustees

Board of Trustees Policy on Public Issues

(adopted October 6, 1981)

The debates over educational testing are part of the democratic process in the United States. The Board of Trustees of Educational Testing Service has participated vigorously in these debates and has learned from them. As Trustees of a publicly chartered body, the Board recognizes its obligation to serve and be responsive to the worthy purposes which led to the establishment of Educational Testing Service thirty-four years ago.

It is the Board of Trustees' intent in establishing this Policy to make clear the corporate position of Educational Testing Service on matters which, in recent years, have been the subject of public debate. In taking this action, the Board is speaking only for Educational Testing Service for which, by law, it is responsible and accountable.

The Board of Trustees reaffirms its position that the public interest is best served by voluntary self-regulation in the educational testing field, subject to public and media scrutiny, rather than by governmental regulation administered by state or federal bureaucracies. It is our position that governmental regulation should be used as a last resort only when responsible educational bodies fail to respond reasonably to legitimate grievances of the public.

By official resolution voted on October 6, 1981, the Board of Trustees of Educational Testing Service formalized its support of the principle of openness in admissions testing. Specifically, the Board supports the recent actions of the College Board, the Graduate Management Admission Council, the Graduate Record Examinations Board, the Law School Admission Council and the TOEFL Policy Council on release of tests and test answers. In taking this position, the Board of Trustees incorporates those actions as policy for Educational Testing Service in the programs it conducts for the respective boards. We commit ETS to advance further, in concert with our contracting boards, the principle of openness in testing.

Openness in testing, however, involves more than test disclosure. The Board of Trustees, acting on behalf of Educational Testing Service, therefore takes the following actions:

- Adopts as policy for Educational Testing Service the "Public Interest Principles" (attached), originally proposed for public discussion on December 30, 1979.

- Approves as policy the "ETS Standards for Quality and Fairness" (attached).* These standards are fully consistent with and indeed, in many respects, exceed the current testing and measurement standards of the American Psychological Association, the American Educational Research Association and the National Council on Measurement in Education.

- Endorses the process for systematic auditing by ETS of all testing programs it conducts to assure adherence to the "ETS Standards for Quality and Fairness." Such audits are performed on a cycle so that each testing program is comprehensively reviewed at least once every three years, with particular attention to provisions in the standards to promote proper test use and to prevent test bias.

- Directs the Trustee Committee on Public Responsibility to prepare annually an accountability report, to be released publicly, which is responsive to public concerns and evaluates actions taken by Educational Testing Service in accordance with the "ETS Standards for Quality and Fairness." In carrying out this responsibility, the Committee will be assisted by a Visiting Committee of persons from outside of ETS who are knowledgeable in the areas addressed by the "ETS Standards for Quality and Fairness."

- Approves establishment of a uniform complaint procedure for students and other test-takers to provide ready access to Educational Testing Service for student grievances about programs and services provided by ETS. A Student Concerns Coordinator will be designated in the President's Office to support this procedure, consult with student organizations and otherwise promote communication with students. The Board recognizes that ETS serves individuals as well as institutions and will endeavor, in full collaboration with its contracting boards, to resolve legitimate grievances about programs and services provided by ETS.

---

* Formerly only administrative guidelines entitled, "Principles, Policies and Procedural Guidelines Regarding ETS Products and Services."

The Board of Trustees further encourages openness in testing for areas other than admissions testing, so that all test-takers will have full and open access to information about the purpose and content of the tests they take, the method of scoring, and the proper uses and limitations of the tests. ETS will work collaboratively with contracting boards and agencies to promote this objective.

The Board of Trustees further supports wide dissemination to schools, institutions of higher education and community agencies of new materials developed to help students and other potential test-takers understand the variety of tests developed by Educational Testing Service. These will include audiovisual materials, videotapes for use by public and cable television, and handbooks to be prepared for publication by the College Board and by Educational Testing Service. The aim of these materials is to make readily available to all students, regardless of economic status, high quality information about tests developed by Educational Testing Service.

It is the Board of Trustees' intent also to broaden access to test and research data which are under the control of Educational Testing Service. The Trustee Committee on Research and Development is developing policy recommendations to increase the availability of this information. The Board expects to act upon these recommendations at its April 1982 meeting. Further, ETS will work collaboratively with its contracting boards to facilitate expanded access to the test and research data which they control, in accordance with policies of the contracting boards. At the same time, the Board of Trustees reaffirms its long-standing commitment to protect the individual privacy of our test-takers in the release and use of test and research data.

Finally, the ETS Board of Trustees herewith urges all profit and nonprofit testing corporations in the United States to join with it in voluntarily developing an industry-wide "Code of Fair Testing" for the development and administration of educational testing nationwide. We recommend that this Code be based on the new standards being developed by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. We further propose that a panel be appointed jointly by these three associations to monitor the industry-wide "Code of Fair Testing" and we commit ETS to cooperate fully with such a panel. Thus, in addition to the actions we have taken on behalf of Educational Testing Service, we offer to join other testing corporations in a national effort to promote openness in testing.

Attachments

Appendix B

PUBLIC INTEREST PRINCIPLES
for the Design and Use of Admissions Testing Programs

Adopted by the Board of Trustees • October 1981

Educational Testing Service • Princeton, New Jersey

PUBLIC INTEREST PRINCIPLES

for the Design and Use of Admissions Testing Programs

By resolution dated October 6, 1981, the Board of Trustees of Educational Testing Service formally adopted as policy for ETS these Public Interest Principles for the Design and Use of Admissions Testing Programs. These principles were originally proposed for public discussion on December 30, 1979, by the leaders of Educational Testing Service and four other organizations responsible for major national admissions testing programs. They address concerns that have been raised about the design and use of standardized tests in admission to higher education, e.g., public access to test questions and answers, verification of scoring procedures, and appropriate use of the information derived from testing programs. Educational Testing Service strongly supports these principles and is committed to working with sponsoring groups to implement the principles in the admissions testing programs with which it is associated.

Principles

A number of the principles enumerated below have been cornerstones of most testing programs for some years. We believe it is important, however, to reaffirm them here to provide a fuller view of our beliefs and our expectations for the future.

1 We recognize the legitimate interest of the public in knowing what the tests contain and their efficacy in performing their intended functions. Therefore, we will implement the principle of publication of test content to a degree limited only by reasonable safeguards of efficiency, cost, quality, and the educational impact of the programs.

2 We fully support the principle of equity and we will continue to maintain and strengthen credible procedures for detecting bias and eliminating it from the content of the tests, while making such procedures visible to the public.

3 We recognize the need for routine procedures that allow the test taker to arrange for verification of the accuracy of the procedures determining the score attributed to him or her.

4 We believe that tests should be readily available to all individuals, regardless of conditions such as physical handicap or religious beliefs that may prevent the taking of exams under circumstances that meet the convenience of the majority.

5 We recognize that tests, together with the procedures for scoring them and reporting the results, should be designed to provide test takers with as much useful information as may be feasible about the specifics of their performance on the tests.

6 We reaffirm the right of individuals and institutions to privacy with regard to information by and about them, which should be safeguarded from unauthorized disclosure.

7 We recognize the need to formulate, maintain, and publish widely principles of appropriate use of scores and other test information derived from testing programs and to be alert to and actively discourage misuse.

8 We recognize that both the institutions making use of test scores and the test takers themselves should have mechanisms through which to express their legitimate interests concerning the design and operation of testing programs and the use of information derived from them.

Operational Elements

The separately constituted and governed groups sponsoring testing programs may choose to implement these principles in different ways. This probable diversity stems from differences in the nature and purposes of the tests in the several programs and from the specifics of their structure and operation. Examples of possible approaches include the following:

1 Each prospective examinee should be able to receive a full-length sample of each test, similar to the one he or she will take, with the intended answers and with instructions for self-administration and self-scoring.

2 For tests given to a sufficient number of students annually to support the cost, at least one operational form of the test should be published periodically, in addition to the regular sample. A specific schedule of publication should be designated for each program.

3 Non-technical information about the testing program should be furnished routinely to test takers, users, and the general public. It should include a description of what each test measures, the error of measurement, how the scores are intended to be used, and a summary of the validity of the scores for the intended uses.

4 A technical publication should provide information on the same topics in sufficient depth to permit professionals in the field to assess the evidence and the accuracy of the non-technical summary.

5 Studies of the use of the test by professionals other than those in the sponsoring or administering agency should be actively encouraged and facilitated by provision of the necessary data with safeguards for individual privacy. The results of those studies should be published in regular journals and also incorporated in the technical and non-technical publications.

6 The test sponsor should ensure that operational forms of the test are independently reviewed before they are given. The review should include the appropriateness of the content of the test and in particular should seek to detect and remove potential racial, cultural or sex bias or other influences extrinsic to the characteristics, skills or knowledge to be measured. The review should also determine that the operational form is fairly represented by the sample test already distributed.

7 Test takers should have the right to question the accuracy of scoring, administrative procedures, specific questions in a test, or allegations of irregularities in test administrations. Current procedures to deal with this right should be reviewed and modified if necessary to ensure a fair and prompt response.

We hope communication of these principles and operational guidelines leads to greater understanding and constructive dialogue about the important issues surrounding testing. We stand ready to work with all interested groups in discussion of the policies and improvement of the procedures under which testing programs are conducted. ■

0151303 • Y101P3 • 292211 • Printed in U.S.A.

Appendix C

ETS STANDARDS FOR QUALITY AND FAIRNESS

Adopted by the Board of Trustees • October 1981

Educational Testing Service • Princeton, New Jersey

PREFACE

By resolution dated October 6, 1981, the Board of Trustees of Educational Testing Service provided for standards for quality and fairness applicable to all ETS activities.

BE IT HEREBY RESOLVED that the Board of Trustees of the Educational Testing Service

1. Adopts as ETS policy and approves as *ETS Standards for Quality and Fairness* the principles and policies previously set forth by ETS management in its publication entitled *Principles, Policies and Procedural Guidelines Regarding ETS Products and Services;*
2. Endorses the procedural guidelines developed and published by management to implement these principles and policies;
3. Directs management to continue its program of monitoring adherence to the procedural guidelines across ETS programs and services, to revise the guidelines as appropriate, and to make an annual report to the Board regarding such efforts;
4. Authorizes management to provide for alternative procedures which can accomplish or hold promise of accomplishing the principles and policies included in the Standards and to make provision for small, new or specialized programs that have special needs; and
5. Directs the Trustee Committee on Public Responsibility to prepare annually an accountability report, to be released publicly, which evaluates actions taken by ETS in accordance with the *ETS Standards for Quality and Fairness*.

The Standards are designed to ensure that ETS products and services meet demonstrable criteria with respect to seven areas of basic importance: Accountability, Confidentiality of Data, Product Accuracy and Timeliness, Research and Development, Tests and Measurement, Test Use, and Technical Assistance, Advice, and Instruction. The first three sections of the Standards deal with issues that relate to all ETS activities: the responsibilities of ETS to those affected by its activities; the rights to and limitations on access to data collected by ETS; and the control of quality and performance according to commitments. The remaining sections concern issues relating to ETS's main endeavors: Research and Development, Tests and Measurement, Test Use, and Technical Assistance, Advice, and Instruction.

The Standards are drawn from particular circumstances and needs at ETS and reflect its objectives. Because of their origin and purposes, the Standards may not be useful to organizations whose practices, programs, or services differ from those of ETS.

The Standards are implemented by ETS management through procedural guidelines which provide more detailed and specific guidance for ETS's diverse programs and services. These procedural guidelines are reviewed and revised from time to time by management as new or alternative approaches are developed. Management also reviews special applications that are necessary to meet the needs of particular programs.

ETS believes that the Standards contribute significantly to the quality and utility of its products and services for clients and consumers. ETS does not have complete responsibility or authority, of course, to determine how the Standards will be implemented in activities for which policy is substantially established by a sponsoring group or institution other than ETS. ETS, however, encourages and assists those groups and institutions to implement the Standards as their activities are related to ETS products and services.

ETS reviews all its programs regularly to assess compliance with the Standards as part of its quality assurance effort. To assist the Trustees to ensure that the Standards are interpreted and applied according to their spirit and purpose, a Visiting Committee of persons outside ETS who are knowledgeable in the specialized aspects of the Standards has been authorized by the Board to review periodically the ETS assessment process. Beginning in 1982, the Committee on Public Responsibility of the ETS Board of Trustees will issue a public report annually on how the organization is meeting these Standards.

These efforts reflect ETS's determination to hold itself accountable to high standards of performance in order to safeguard the public interest and the interest of individuals using its products and services.

October 1981

Gregory R. Anrig

Gregory R. Anrig
President

CONTENTS

Page

Preface . . . iii

Accountability . . . 1

Confidentiality of Data . . . 4

Product Accuracy and Timeliness . . . 7

Research and Development . . . 9

Tests and Measurement—Technical Quality of Tests . . . 12
- Test Development . . . 13
- Test Administration . . . 16
- Test Reliability . . . 19
- Scale Definition . . . 20
- Equating . . . 22
- Score Interpretation . . . 23
- Test Validity . . . 25

Test Use . . . 27

Technical Assistance, Advice, and Instruction . . . 29

Glossary . . . 31

ACCOUNTABILITY

Principle

ETS acknowledges responsibility for the effective stewardship of its resources to the New York Board of Regents which has issued its corporate charter, to the governing boards that sponsor and set policy for programs or services in which ETS products or services are used; to the individuals and committees that advise ETS with respect to appropriate policy for its programs; to the institutions and agencies that use ETS products and services; to persons who take ETS tests (and parents or guardians of minor persons), submit data for use by ETS or for distribution to others, or participate in research and development projects conducted by ETS; and to the professional associations that are concerned with educational and psychological measurement and research.

Policies

A. ETS will furnish appropriate information to those to whom it is responsible so they may make informed, independent judgments as to the effectiveness with which ETS exercises its stewardship.

B. ETS will seek, consider and, as appropriate, act on the views of those who sponsor, use or are affected by ETS programs and services.

C. ETS will seek to obtain advice on its activities and policies from qualified men and women who are not employed or retained on a regular basis by ETS and who are drawn from appropriate professional disciplines, major philosophies and points of view, different geographic regions, and the major ethnic groups within the relevant population.

D. ETS will support the activities of professional associations with respect to developing and implementing professional standards or codes, making available the results of current work, and fostering peer review of its activities.

Procedural Guidelines

1. Information should be provided to sponsoring organizations with which ETS has contractual relationships in a form that permits evaluation of ETS services in terms of:

   a. quality;
   b. timeliness;
   c. costs; and
   d. responsiveness to legitimate comments or criticisms.

2 2. Procedures should be established to facilitate communication with sponsors by:

   a. meeting at least annually to provide information and to receive comments on matters affecting the operations with which they are concerned;

   b. defining a mutually agreeable process to be used to transmit comments from sponsors or others and a time period within which the evaluation of comments will be completed and reports of actions to be taken by ETS can be expected; and

   c. making available periodic opportunities for sponsors to express opinions, judgments and counsel concerning their activities or programs directly to ETS officers not normally responsible for such activities or programs.

3. Procedures should be established for making available technical and other information about ETS products and services to users so that they may evaluate the appropriate use of the product or service and communicate comments or criticisms to ETS.

4. Procedures should be established to communicate with, or provide information to, persons who use or take ETS tests, who submit data for use by ETS or for distribution to others, or who participate in research and development projects conducted by ETS. This information should be communicated by ETS or the sponsor in such a way that these persons may understand their participation with respect to:

   a. the identity and scope of the sponsor's responsibility;

   b. the nature of the product, service or research by which they are affected;

   c. the way in which the product, service or research will likely be used by educational institutions or others; and

   d. the channels that have been established for addressing comments or criticisms to ETS or to the sponsor and response thereto.

5. Organizational and program financial information should be recorded, processed and reported in accordance with generally accepted accounting principles and under appropriate safeguards to insure accuracy.

6. An annual report that provides information about organizational activities and finances should be published by ETS on a regular basis and made available to any person on request. Program and project reports, including program financial information, should be made available in a manner consistent with contractual understandings.

7. Requests for information that is not included in an existing publication should be considered by the appropriate sponsor and by ETS. If its disclosure is consistent with applicable law, with ETS and sponsor policy, and with contractual obligations governing confidential or proprietary information, the information should be provided. If complying with a request for information results in a cost to ETS or a sponsor or affects the normal schedule of fulfilling ETS's responsibilities, ETS may provide the requested information in a reasonable period of time and at an appropriate price for the services rendered. Procedures should be established, as appropriate, to facilitate responses to these requests.

8. Changes in federal statutes, regulations and case law that affect research and development, testing programs, or advisory and instructional services should be monitored to ensure that ETS activities and operations are in compliance as relevant federal laws or rules change. Changes in other statutes, regulations and case law should be evaluated, as appropriate, for the same purpose.

9. All proposed new ETS activities should be reviewed by counsel for compliance with applicable federal law and state law, as appropriate. ETS officers and staff should direct the attention of legal counsel to matters that might affect ETS compliance.

10. Advice should be sought, where appropriate, from men and women drawn from diverse backgrounds, interests and experience (e.g., appropriate professional disciplines, major philosophies and points of view, various geographic regions, and major ethnic, handicapped and other relevant subgroups of the population of interest) who are qualified to make a contribution to the direction and substance of ETS programs and who are not employed or retained on a regular basis by ETS.

11. Individuals who become members of an ETS external advisory, review or evaluation committee should be informed about the results of the committee's work in a reasonable period of time.

12. A reasonable accommodation should be made with respect to the professional responsibilities of the staff in order to permit staff members to attend professional meetings, to contribute to the development of professional standards or codes, to participate in and benefit from the dissemination of information on subjects of professional interest, and to stay abreast of current concerns and accomplishments in related fields.

13. ETS should have effective procedures for peer review whenever it will contribute substantially to the quality of ETS work.

14. ETS should have effective and equitable procedures for handling questions of score authenticity arising in connection with the administration of tests.

4

CONFIDENTIALITY

Principle

ETS recognizes the right of individuals and institutions to privacy with regard to information supplied by and about them that may be stored in data or research files held by ETS and the concomitant responsibility to safeguard information in its files from unauthorized disclosure.

Policies

A. ETS will ask individuals to provide information about themselves only if it is potentially useful to those individuals, necessary to facilitate processing of data or serves the public interest in improving understanding of human performance. Insofar as possible, individuals should be informed of the purpose for which the information is requested.

B. The right of individuals to privacy regarding information about them that may be stored in the data or research files held by ETS extends both to processed information, such as scores based on test-item responses, and the raw data on which the processed information is based.

C. ETS will protect the confidentiality of data supplied by institutions or agencies about themselves, and so identified, to the extent that such confidentiality does not conflict with ETS's obligations to individuals.

D. ETS will not collect or maintain in its data or research files any critical information that in its judgment cannot be protected adequately from improper disclosure.

E. ETS will encourage the organizations with which it works to adopt policies and procedures that adequately protect the confidentiality of the data transferred by ETS to those organizations.

Procedural Guidelines

1. Information about an individual, which has been identified as such, may not be released by ETS to organizations other than those for which the information was collected without the consent of that individual. A written exception may be made in the case of research studies during which the Committee on Prior Review of Research has determined that release of the data serves a public need, that there is no satisfactory and reasonable alternative way of obtaining the information, that the recipient researcher will use the data in appropriate ways and that there are adequate assurances of confidentiality.

2. Information about an institution, which has been identified as such, may be released from ETS only in a manner consistent with a prior agreement or with the consent of the institution or with the approval of the cognizant ETS officer and representative of the appropriate sponsor (if any).

3. An individual should be able, on payment of a reasonable fee, to authorize the disclosure of information about himself or herself from program data files held by ETS to any appropriate recipient, provided that such authorization is in writing and that disclosure is not inconsistent with other ETS or sponsor policies and does not violate the privacy of other individuals. Identification of the requester, through signature and data file number, or other appropriate method, should be required before any such information is released.

4. In an emergency and when it is to the benefit of the individual, an authorization by telegram or telephone for the release of personal data should be acceptable, provided that such authorization includes adequate identifying information and that such release is not inconsistent with other ETS or sponsor policies. By prior agreement with the individual, authorization by a designated agency or institution should also be acceptable. In such instances, the individual should be informed that the disclosure has taken place.

5. If an individual is not competent because of illness or other considerations, information about that individual may be released from data files only with the consent of the individual's parent or legally appointed guardian.

6. Unless the access to confidential data can be safeguarded, ETS should not participate in any time-sharing network, data bank, or other electronic data processing or storage system involving units outside ETS.

7. On submission of appropriate identifying information and payment of a reasonable fee, an individual should be able to obtain information about himself or herself in ETS-held data files for the following purposes: to ascertain the accuracy of personal or biographical data and to request verification, within a reasonable period of time, of test scores or other processed information from tests, questionnaires, or school records, provided such release is consistent with sponsor policies.

8. Procedures should be developed for systematically eliminating from data files information that is judged to be out of date and, hence, of minimal value.

9. Information from ETS-held program data files provided by individuals for a designated purpose should not be used or released for another purpose (such as a validity study or research project) without the individual's consent except when used or released in a form that cannot be identified with the individual.

6

10. ETS should refuse to provide personally identifiable information except in accordance with these guidelines unless served with a subpoena or other court order. In that event, ETS should make appropriate efforts to quash or narrow the subpoena or order or to obtain a protective order to minimize the exposure of personally identifiable information.

11. At the time information is collected and to whatever extent practical, programs should inform individuals of the conditions surrounding the release and confidentiality of the information about them.

12. Individuals should be identified in ETS research files only by code numbers. Information linking the code numbers to names should be kept in a secure location only as long as necessary for purposes such as follow-up studies or collating new data, after which the names should be destroyed.

13. Every organization with which ETS works should be informed of the confidential nature of any data transferred by ETS to that organization or collected by the organization on behalf of ETS so that appropriate procedures can be employed by the recipient organization to protect the confidentiality of such data.

PRODUCT ACCURACY AND TIMELINESS

Principle

The accuracy of ETS's principal products and the timeliness with which they are made available are important parts of the responsibility ETS has undertaken with respect to its sponsors and the diverse public it serves.

Policies

A. ETS will establish standards of accuracy and timeliness with respect to each principal product.

B. ETS will use quality controls that are adequate to assure that its standards of accuracy and timeliness are met.

C. ETS will make realistic delivery commitments and reasonable efforts to meet those commitments.

D. ETS will sacrifice the timeliness of the delivery of information if the desired accuracy of that information is substantially in question.

E. ETS will seek to inform those adversely affected if, subsequent to its release, information has been found not to meet ETS standards of accuracy.

F. ETS will seek to inform those adversely affected if there is a probability that there will be substantial departure from ETS standards of timeliness with respect to a principal product.

Procedural Guidelines

1. Principal products should be identified and a standard of accuracy using units of measurement appropriate to the type of product should be established for each.

2. When appropriate, quality control should include an adequate and independent recomputation and a visual reexamination of ETS-processed information based on an appropriate sample of cases sufficient to identify errors within the limits of the applicable standards of accuracy.

8 3. When the computational nature of the information is such that it is impossible or impractical to determine the accuracy of the information by independent recomputation, staff members who are technically competent to do so should assess its "reasonableness" as a part of quality control.

4. There should be a quality inspection of intermediate products when:

   a. the accuracy of variable information (e.g., parameter data, algorithms), verified by independent recomputation or assessment, influences the nature of an ETS process or computation and is critical to the process that generates an ETS principal product; or

   b. detection and correction of errors would facilitate meeting the delivery commitment on the principal product; or

   c. the nature of the principal product is such that it is impossible or impractical to determine the accuracy of the information by independent computation using the source data.

5. Quality control for principal products that do not contain processed information (such as bulletins of information or test books) should include inspection of a sample prior to release of the product. If the product is released from an outside vendor (e.g., outside publisher) or a sponsor's agent, quality control should include inspection of those components of the principal product that contain critical information on ETS-provided services.

6. Quality control of information given in letter or telephone responses should include a periodic audit of a sample.

7. Failure to meet standards of accuracy and timeliness should be reported to a designated ETS staff member for resolution.

8. A principal product that does not meet established standards of accuracy should not be released until appropriate corrective action is taken, unless release would be for the benefit of the score recipient and users and permission to release is given by the cognizant ETS officer.

9. If an error is found in critical information already released by ETS, the correct information should be promptly distributed.

10. Process control methods (e.g., a predefined schedule including a delivery date and contingency procedures for dealing with volume surge) should be established for the production of each principal product to help assure its delivery by the scheduled delivery date.

11. If it is likely that there will be a substantial departure from ETS standards of timeliness with respect to a principal product, those who would be adversely affected should be so notified.

RESEARCH AND DEVELOPMENT

Principle

A continuing program of research and development conducted in compliance with professional standards with respect to quality and ethical procedures is necessary to maintain the high quality and social utility of ETS contributions to education. This includes basic inquiry to increase understanding of educational processes and human development; evaluative and applied research in response to the needs of the educational community; and research and development to improve ETS products and services. Publication of the results of significant ETS research is of benefit to ETS and the profession because it permits others to use, build upon or improve ETS work.

Policies

A. ETS will devote appropriate research efforts to improving education through the discovery and conceptual integration of new principles and understanding. This research will be aimed at extending knowledge of the learner and learning processes, of learning environments and educational treatments, of educational institutions and of the interacting factors that influence human development.

B. ETS will devote appropriate research efforts to the improvement of the technical quality of ETS products and services. Among the important issues addressed by this research will be problems of test development, reliability, equating, validity, and meaningfulness of interpretation.

C. ETS will devote appropriate research and development efforts to the identification of needs of the educational community and to the creation, improvement and evaluation of instruments, systems and programs of service that meet these needs.

D. ETS will conduct its research under appropriate procedures that protect the rights of privacy and confidentiality of human subjects or respondents.

E. ETS will follow procedures to insure that ETS research is of high quality. Standards of quality in research refer to such matters as the identification of relevant data, the choice of suitable methods of collecting and analyzing data, the logic and objectivity of analysis and interpretation, the exploration of relationships between research problems and findings, on the one hand, and existing knowledge, theories and methodologies on the other, and the thoroughness and care of project planning and management.

F. ETS will undertake research only if its potential benefits outweigh the inconveniences of or risks to the subjects or respondents who are involved.

G. ETS will encourage the dissemination of full accounts of ETS research in the usual professional forums and will provide internal means by which the results of ETS research can be published.

Procedural Guidelines

1. To maintain the quality of operational programs, ETS should engage in the following activities:
    a. study and research on the test development process, including systematic development and evaluation of new item types and approaches;
    b. studies to determine the sources of significant differential performance of sex, ethnic, handicapped, and other relevant subgroups on ETS tests;
    c. periodic evaluation of current approaches to aptitude and achievement measurement to determine fairness, validity and appropriateness for significant subgroups such as minorities and women;
    d. research related to reliability theory and practice, including methods of determining the reliability of classification decisions;
    e. study of the equating methods presently in use and development of improved methods as limitations in the applicability of the present methods are observed; and
    f. research to advance measurement techniques and selection and classification models relevant to fairness and validity.

2. Research projects should be undertaken in such areas as learning and cognition, personality and social influence, teacher behavior and instructional processes, socialization and human development, and the economics and sociology of education as a means of improving educational policies and practices.

3. Efforts should be made to develop instruments and programs of service in areas such as measurement, institutional and program assessment and evaluation, instruction, guidance, financial aid, certification and licensing, and technology that would be of educational and social utility.

4. Proposals for research to be conducted by ETS and involving human subjects or respondents should be considered by the Committee on Prior Review of Research, under its procedures for review, to verify that proper arrangements have been made for protection of the welfare and rights of human subjects.

5. Researchers should not conduct research projects without the consent of subjects and respondents. In the case of young children, the consent of parents or a legal guardian, or of appropriate institutional representatives, should be obtained.

6. Each research proposal should be reviewed by one or more persons who are competent in the field within which the proposal falls. They should be satisfied that professional standards of quality and ethical conduct are met.

7. Identifiable data should be released from ETS to researchers other than those who originally conducted the research only when one of two conditions have been met:

    a. Consent to do so has been given by or on behalf of the subjects or respondents or by those who have given consent on their behalf; or

    b. the Committee on Prior Review of Research has determined that release of the data serves a public need, that there is no satisfactory and reasonable alternative way of obtaining the information, that the recipient researcher will use the data in appropriate ways and that there are adequate assurances of confidentiality.

8. After the data-collection phase of a research project has been completed, subjects should not be expected to provide additional data for a follow-up study unless such participation was part of their original agreement to serve as subjects, or their consent for follow-up is obtained or the follow-up study has been approved by the ETS Committee on Prior Review of Research.

9. The results of measures of performance based on experimental situations or tests, the interpretation of which is, therefore, tentative, and whose applied use is not yet supportable, should not be reported to subjects, or to the institutions providing the subjects, unless there is relatively little danger of misinterpretation or misuse of the information that would be harmful to those individuals or institutions, or unless the use is part of a feasibility study or experimental condition. Stipulations regarding nonissuance of such reports should be made to participants in advance of the data collection.

10. The results of each research project undertaken with respect to a particular ETS program or service should be available for dissemination unless a specific need to restrict publication to protect confidentiality or for other program purposes is identified prior to the beginning of the project and made known to the appropriate individuals.

11. The contracts under which research is undertaken for agencies or institutions outside ETS should permit publication of the results of the research unless a specific need to protect the research results is identified prior to the beginning of the research and made known to the appropriate individuals.

12

TESTS AND MEASUREMENT—TECHNICAL QUALITY OF TESTS

This section which deals with ETS testing activities is divided into seven subsections that are devoted to test development, test administration, reliability, scale definition, equating, score interpretation, and validity.

Principle

High standards of quality and fairness in constructing, administering, reporting, interpreting and evaluating ETS tests are central to ETS's capability to function effectively as an educational service and research organization.

Policies

A. ETS will strive to develop tests in which the attributes measured, procedures followed, and criteria used will be unbiased with regard to a heterogeneous group of examinees and appropriate to the use for which the test is designed.

B. ETS will establish standards for test-administration processes that minimize variations in test performance due to circumstances or conditions not relevant to the attributes being measured.

C. ETS will establish for its tests a high degree of reliability (accuracy of measurement), consistent with the requirements and the purposes of the test.

D. ETS will develop scales for reporting scores in a rational fashion, consistent with the requirements and the purposes of the test.

E. ETS will provide equating systems, when appropriate, for the perpetuation of scales for reporting scores at the highest level of precision practicable.

F. ETS will make available to sponsors, institutional or agency users, and examinees data for interpreting scores on ETS tests that foster appropriate use of those scores.

G. Recognizing that test validation is a responsibility of both test users and test developers, ETS will encourage and assist test users in their validation efforts and will itself make available tests that are designed to meet professionally acceptable standards of validity provided the use of such tests is consistent with the primary purposes for which the tests were developed.

H. ETS will adhere to appropriate professional standards, such as those published in *Standards of Educational and Psychological Tests* and *Principles for the Validation and Use of Personnel Selection Procedures*.

Procedural Guidelines: Test Development

1. Policy and substantive contributions to the test development process should be obtained from qualified men and women who are not on the full-time staff of ETS and who are drawn from diverse backgrounds and appropriate specialties within professional fields (e.g., various kinds of institutions and programs, relevant philosophies and points of view, and major ethnic, handicapped and other relevant subgroups of the population).

2. Appropriate background information for use in the development of a test should be documented at appropriate stages in the development process and include:

   a. the purpose for which the test is intended to be used;

   b. the nature of the population that will take the test;

   c. the relevant procedural, financial or time constraints that will influence the available test development methods and their likely outcomes;

   d. for achievement tests, the kinds of curricula for which the test is designed; and

   e. for job-related tests, the elements in training or employment that are related to performance on the job.

3. For each test, specifications should be developed and reviewed by a process that provides information from the following perspectives:

   a. Content and Skills—specifications should include the psychological, educational, or other domains to be sampled; the relative weight to be given to each domain; the appropriate level of proficiency to be required within each domain; a balance with respect to curricular differences.

   b. Test and Item Format—specifications should include the item (question) types that are most clearly related to content or skills to be measured; the appropriate level of language or reading; requirements regarding clear and comprehensive directions and sample items or the need for a sample test; and whether free-response, multiple-choice or other machine-scorable formats can be used.

   c. Psychometric—specifications should include the level of difficulty of the test; the distribution of item difficulties (when pretested items are used); guidelines for evaluating the homogeneity among items within a test and the relationship between subtests or tests; equating requirements; number of items and time allotted.

   d. Sensitivity—specifications for tests should require material reflecting the cultural background and contributions of women, minorities, and other subgroups; specifications should also require a balance of positive connotations if negative connotations are made in any references to these groups.

4. Except for tests designed to measure rate of performance, the number of items in a test that has a specified time limit should be chosen so that time is not a decisive factor in performance, at least for the large majority of examinees.

5. Subject matter and measurement specialists familiar with the purpose of the test and with the characteristics of the intended population should review the test items for accuracy, content appropriateness and the adequacy with which the items sample the domain.

6. The individual items in a test should meet appropriate technical standards such as those contained in the manuals for item writers used in the test development area.

7. Individual test items, and the test as a whole, should be reviewed to eliminate language, symbols or content which are generally considered potentially offensive, inappropriate for major subgroups of the test-taking population or serving to perpetuate any negative attitude which may be conveyed toward these subgroups. No item in any test should include words, phrases or description that is generally regarded as biased, sexist or racist (e.g., demeaning modifiers and stereotypes).

8. The items in a test should be reviewed by editorial specialists for clarity, accuracy, consistency, and, when appropriate, for conformity with standard editorial style.

9. Tests should contain clear and complete directions. Enough sample problems should be provided in test-program publications so that the examinee can understand the nature of the task and the test-taking procedures. Where there is a need to provide a general orientation to testing, as when testing young children, practice tests—included either in descriptive material or at the time of test administration—should be used.

10. The typography, directions, and arrangement of items in the test booklet should facilitate the task of test takers. When appropriate, tests should be made available to handicapped individuals, such as sight-deficient candidates, through tapes, readers or special printing.

11. Methods should be employed to evaluate the appropriateness of items before their operational use in a program or before the reporting of scores. Appropriate methods include pretesting, preliminary item analysis (using the first operational use of items as an opportunity to identify inadequate items) or careful review of the results of administering similar items to a similar population. In assessing the appropriateness of items before their operational use, efforts should be made to include representative samples of the operational test-taking population.

12. The operational use of each test should be followed by systematic item analyses using appropriate criteria and by test analyses. These analyses should include reliability, intercorrelations of sections or parts, and speededness.

13. Studies relating item performance to subgroups should be carried out for new or substantially revised tests when there are adequate data concerning sufficient samples of large subgroups whose education and experience may be different from the majority of examinees.

14. The specifications for tests in ongoing programs should be reviewed for relevance and appropriateness before each new form is created. ETS staff and advisers should consider whether changes in the field, discipline or curricula require a revision of the specifications.

15. When major changes are made in test specifications, consideration should be given to the implications of such changes for score comparability and whether it is necessary to change the test name or otherwise communicate to those who interpret test scores that comparisons with earlier tests may be inappropriate.

16. When test forms are used for a number of years in a program, they should be reviewed periodically for their appropriateness. The frequency of such review should be determined by the amount of change occurring in the population of test takers or the subject matter domain. Test forms that are found to be outdated should be revised or withdrawn from use.

Procedural Guidelines: Test Administration

1. Information should be made available to prospective examinees and (in some programs, to parents or guardians as well) in advance of the test administration with respect to the following, as appropriate:

   a. the purpose of the test and what it measures;

   b. the nature of the test items (including samples of typical item types);

   c. the relevant instructions for taking the test, including instructions for guessing, changing answers, and strategy involving speed and accuracy in taking the test;

   d. identification requirements and the consequences of not having identification;

   e. the consequences of misconduct by the test taker;

   f. background and experience relevant to test performance;

   g. the location of test centers, the test dates and special testing arrangements that can be made;

   h. the procedures for registering for the test and changing the centers;

   i. the structure of test fees and fee waivers;

   j. special arrangements available for administering tests to handicapped individuals;

   k. the reporting of scores;

   l. procedures for canceling test scores by the candidate and reasons why ETS or the sponsor of the test might cancel scores; and

   m. the procedures for registering complaints.

2. Program publications should be reviewed for language or descriptions generally regarded as biased and offensive. For example, the exclusive use of masculine pronouns should be avoided as should the implication that all persons in a given category (for instance, examinees, supervisors, counselors, or teachers) are either females or males (unless, of course, the category is logically restricted to members of a single sex). Illustrations, examples and practice items in test-information publications should represent males, females, minority and majority groups, and individuals in ways that indicate respect and awareness of valuable contributions.

3. The facilities at which tests are administered should be places that are convenient for the majority of examinees, nonsegregated and comfortable. At least portions of those facilities should be accessible to and responsive to the needs of handicapped individuals.

4. ETS should enlist test center supervisors and staff with demonstrated sensitivity to the anticipated sex and ethnic composition of the examinee group, based on prior experience. When appropriate, persons affiliated with institutions attended by significant numbers of those examinees should be included. Minority group supervisors and/or proctors should be employed, and test sites should be located in minority communities whenever appropriate and feasible.

5. Test center supervisors and staff should be familiar with the procedures for administering a standardized test and should be provided with a description of the testing program, a description of the candidate population, and specific instructions for administering the test. Instructions concern such subjects as the duties of test supervisors, associate supervisors, and proctors; the receipt, storage and return of test supplies; the admittance of examinees to the testing rooms; the distribution of test materials; procedures to be followed in administering tests to handicapped individuals; procedures to be followed in instances of suspected cheating; procedures to be followed in other cases of candidate misconduct; and procedures to be followed in case of emergency.

6. Test performance can be affected by the psychological atmosphere of the testing center. Test supervisors should be informed of this and instructed to take measures to avoid an adverse situation. For example, test supervisors should be instructed, when it is appropriate and feasible, to have minority- as well as majority-group persons, women as well as men, read test directions and to recognize questions from examinees following an impartial procedure.

7. ETS should provide the test center supervisor with directions to be read aloud to examinees before the test begins. These directions should include information relating to: procedures for marking answer sheets, timing of test sections, strategies for guessing, time and duration of test breaks and examinees' use of unauthorized aids. Test supervisors should check to see that examinees understand their task and the procedures to be followed.

8. Reasonable efforts should be made to eliminate opportunities for examinees to attain scores by fraudulent means by stipulating requirements for identification, assigning examinees to seats and requiring appropriate space between seats.

9. Appropriate procedures should be applied after the test administration to identify scores of questionable authenticity, to resolve issues of authenticity and to provide for prompt reporting of questioned scores found to be authentic.

10. A systematic program for observing test administrations should be conducted by trained ETS staff members or other qualified individuals: to review the testing procedures with the test supervisors, to ensure appropriate testing conditions, to ensure adequate maintenance of test security at the test centers and to relay questions and concerns from the field to the appropriate ETS office.

11. Testing programs should have detailed procedures for investigating and resolving examinees' complaints of irregular test administration or score reporting.

12. Comments and suggestions should be solicited from supervisors by such means as the Supervisor's Comment Sheet and meetings of supervisors to provide ETS staff with information to improve future administrations.

13. Supervisors should be required to record and report to ETS information on irregularities (such as mistiming, defective materials, power failures and cheating) so that ETS can evaluate the possible effect of such occurrences on examinees' test performance.

14. An individual who has taken a test should be provided information that will be helpful in interpreting scores on that test.

Procedural Guidelines: Test Reliability

1. When test scores are reported to institutional or agency users or to individual examinees, information about the reliability of the test should be documented and should include:

    a. a reliability coefficient and an overall standard error of measurement (several indices may be provided if more than one method of assessing reliability has been used; alternate-form information should always be provided if available);

    b. standard errors of measurement for score regions if decisions about individuals are made in those score regions and if the overall regions and the overall standard error are judged inappropriate;

    c. the formula(s) used to estimate reliability and/or appropriate references;

    d. a justification of the method(s) used to assess reliability;

    e. a specification of the major sources of measurement error accounted for in the reliability analysis;

    f. a specification of the time interval between testings if alternate-form or test-retest reliability is used;

    g. the number of observations, the mean and standard deviation of the analysis sample (ranges or averages are acceptable in cases where the reliability information is derived from several samples);

    h. speededness data; and

    i. correlations of subscores within the same test or battery of which the test is a part.

2. If reporting any of the reliability information required under Guideline 1 is inappropriate, the reasons should be stated in appropriate program documents and, if possible, alternate information about consistency should be provided.

3. Efforts should be made to provide reliability information in an appropriate form to the examinees to whom the scores are reported.

4. The method(s) used for assessing reliability should:

    a. take into account the most common sources of error generally considered significant for test interpretation (e.g., guessing, instability over time, item and content variation, and rater inconsistency); and

    b. be appropriate to the nature of the test, in order not to seriously over- or underestimate reliability.

Procedural Guidelines: Scale Definition

1. Raw scores on a test or subtest (including percentages of questions answered correctly) should not be reported by ETS for individual examinees or in summary form for groups of examinees except under either of the following circumstances:

    a. when it is anticipated that only one edition of the test will be offered for use in the foreseeable future or it is demonstrated by appropriate empirical procedures that raw scores on all the editions to be compared are interchangeable; or when raw scores on that test edition will not be compared directly with raw scores on another test edition; or

    b. when reported in conjunction with a scaled score and in a context that supports appropriate interpretation, such as when a copy of the test itself is available or when individual or group responses to individual items, depending on whether individual or group performance is being assessed, are available.

2. If a test or test battery yields multiple scores for an individual and scaled scores are to be used directly (e.g., without reference to norms tables) in interpreting performance profiles, the scales should be normatively defined and each should be defined with respect to the same population.

3. When different tests in a program are taken by different examinees whose scores are to be directly compared, the scales for the tests should take into account possible differences among the groups of examinees who take the various tests.

4. Established scales should not be redefined except under compelling circumstances. If a scale is to be substantially redefined, the numerical values should be changed substantially to minimize the possibility of confusion between test results expressed on the revised scale and results expressed on the original scale. An exception to this guideline may appropriately occur if the test in question is one of a set of tests for which a single range of numerical values (e.g., 20-80) is used and the scales for other tests in the set have not been redefined.

5. Scale properties that affect score interpretation and use should be described in program publications available to the examinees and to institutional or agency users.

6. Technical manuals and interpretive publications for institutional or agency score users and examinees should indicate, in language appropriate to the audience, whether a distributively based scale is intended to be normative or nonnormative. If it is intended to be normative, the group should be described.

7. Whenever a normatively defined scale no longer conveys useful normative information, all published descriptions of the scale should be changed accordingly.

8. Program publications should caution score recipients (users and examinees) that scores received on different tests that are reported on scales that are similar in appearance may not be equivalent.

Guidelines 9 through 14 apply only to scales established after guidelines 1-8 were published on August 1, 1977.

9. If a scale is to be distributive, the choice between a normative and nonnormative distributive scale should take into account:
    a. the extent to which normative interpretation with reference to a particular population will be appropriate and useful for all examinees who take the test and for all purposes for which the scores are intended to be used;
    b. the probable time period during which the normative information conveyed by the scores will continue to be descriptively appropriate; and
    c. the feasibility of identifying and testing a suitable group of examinees on which to base a normative scale.

10. The choice between a distributive and nondistributive scale should take into account the use for which the test was intended and to which the test is likely to be put.

11. If a scale is to be defined with reference to standards of performance, the basis for establishing the standards should be determined empirically or rationally rather than arbitrarily.

12. The conventional grade- or age-equivalent score (the grade or age for which a particular score is the average) should not be used to establish the score scale for a test or system of tests. This type of score, as it typically has been derived, should be avoided altogether as a basis for reporting test performance. However, the grade (or age) for which a particular scaled score on a test is the average, referred to here as a "grade (age) level indicator" to distinguish it from the conventional grade-equivalent (age-equivalent) score, may be reported to help in score interpretation, if the practices customarily followed in deriving and presenting grade-equivalent (age-equivalent) scores are modified in accordance with ETS criteria that obviate the technical interpretive problems that grade-equivalent (age-equivalent) scores create.

13. The choice of a scale should take into account the likelihood of confusion with other widely used scales.

14. In establishing the number of distinct scale values to be reported, consideration should be given to the relative importance of the need to avoid erroneous distinctions among individuals (by reporting different scores for individuals whose true scores are the same) and the need to maintain distinctions that, on the average, will be correct (by reporting different scores for groups of individuals whose average true scores are different).

Procedural Guidelines: Equating

1. Adequate equating should precede comparisons of the test performance of two or more individuals or groups on nonidentical items or sets of items such as test offerings in which successive, or alternate, forms are used interchangeably.

2. Statistical methods selected for equating should be used only under circumstances that are consistent with the assumptions under which the methods have been developed.

3. In regular and continuing testing programs that are available to users, integrated, long-range systems of equating the scores on all successive editions of the test should be used and described in technical publications.

4. For those tests that are offered for institutional use (as distinguished from externally administered tests offered in testing programs) of which only a limited number of forms are available, equating of new forms should be based on specially designed studies in which examinees or groups of examinees are selected by an appropriate sampling procedure to take the alternate forms or alternate sequences of forms.

5. When test forms are equated with the use of common (anchor) items, the psychological task of taking those items (represented, for example, by the directions, the context of the items and the speededness of the part of the test in which the items appear) should be the same for all examinees.

6. When the common items used for equating are not representative of the tests being equated, the groups of examinees used for equating should be as nearly as possible equivalent.

7. In the continuing testing programs, statistical checks (e.g., check equating, special scale-stability studies) should be employed to permit regular assessment of the precision of the equating.

Procedural Guidelines: Score Interpretation

1. Effective test use and meaningful score interpretation should be supported and augmented by:

   a. the development of appropriate test norms based on administering tests to samples from a defined population when there is a reasonable expectation that a large proportion of the schools or other units selected for the norms sample will agree to participate; or

   b. a rationally developed system of interpretation shared with score recipients when score interpretation is not developed from normative data.

2. Tests offered for sale and described by ETS as standardized tests (as distinguished from tests offered in testing programs) should have adequate norms or other information for use in interpreting test results.

3. When test norms are developed by administering tests to samples from a defined population, the resulting norms should be representative of any relevant subgroup, including those defined by sex or ethnicity, in proportion to their frequency in the defined population. Such subgroups may be deliberately over-sampled for more precise estimation of the statistical characteristics of the population by procedures that take over-sampling into account. Data on the proportions in the sample and in the population, when available, should be reported in an appropriate technical publication.

4. The report of a special norms study should provide information on:

   a. the sampling design;

   b. the participation rate of institutions or individual respondents in the sample;

   c. characteristics of the participating institutions and individuals;

   d. weighting systems used in preparing norms; and

   e. estimates of sampling variability along with an acknowledgment, when necessary, that such estimates do not take into account biases arising from nonparticipation.

5. When descriptive statistics based on program testing (as distinguished from norms based on special norms studies) are published, the following guidelines should be used:

   a. both table titles and descriptive material should make it clear that the statistics are based on examinees or participating institutions or other using agencies;

   b. the descriptive material should define the nature of the group by identifying the appropriateness of the sample and the factors that relate the background of the group to test performance, and by acknowledging explicitly that the sample is self-selected;

c. when possible, reports should be prepared to show comparisons of data based on program examinees or institutional characteristics with relevant data on variables from other sources; and

d. when information about interpretive data is prepared for different user groups, the presentation, whenever practicable, should be adapted to the needs and background of each group.

6. When norms are developed from program testing, the age, sex and ethnic composition of the program norms group should be described whenever such information about subgroup membership is available.

7. In testing programs, descriptive statistics should be compiled periodically from a sample or entire population in order to monitor the participation and performance of males and females drawn from diverse backgrounds, interests and experience (e.g., major ethnic group, handicapped status and other relevant subgroups of the population of interest).

8. If norms intended for use in the interpretation of individual scores are presented separately for males and females or for members of specific ethnic groups, the rationale should be carefully described. Separate norms may be justified for scores used primarily for guidance when access to the experiences needed to earn a high score is clearly related to subgroup membership and a more direct index of access is not available. The existence of score differences between subgroups does not in itself justify presentation of separate norms.

9. Descriptive statistics prepared separately for subgroups of the relevant test-taking population, but not intended for use in interpreting individual scores, should not be presented in a way that encourages their use for such a purpose.

10. Institutional or agency users and examinees should be informed of the standard error of measurement of a score, and test interpretation materials should point out the limitations of test scores and encourage score users to take into account the possible scores a test taker might achieve on retesting.

11. Statistical data used in score interpretation should be revised annually except when less frequent revision is judged to be appropriate as, for example, when norms are based on special studies. A statement of the period in which the data were collected should be included in any publication that presents the data.

12. Institutional or agency score recipients should be provided with interpretive materials designed to be helpful for using scores in conjunction with other information, setting cutting scores where appropriate, interpreting the scores for special subgroups (e.g., ethnic minorities, males, females, and handicapped students), conducting local normative studies, and developing local interpretative materials.

Procedural Guidelines: Test Validity

1. ETS should provide evidence of the validity of its tests in relation to the principal purposes or intended uses of the tests. One or more of the following may be applicable:
   a. When test scores are to be interpreted in terms of degree of mastery of the knowledge, skills, or abilities of a domain represented by the test, content validation evidence should be provided.

   b. When test scores are to be interpreted in terms of the prediction of future behavior, criterion-related validation evidence should be provided.

   c. When test scores are to be interpreted as a measure of a theoretical construct, construct validation evidence should be provided.

2. Evidence of content validity should be based (a) on a careful determination and analysis of the domain(s) of interest and of the relative importance of topics within the domain, and (b) on a demonstration that the test is an appropriate sample of the knowledge or behavior in the domain(s). A report on evidence of content validity should present descriptions of the procedures employed in the study, including the number and qualifications of experts involved in the analysis of the domain or evaluation of the relevance and appropriateness of the test.

3. Construct validation should be based on: rational and empirical analyses of processes underlying performance on the test in question including, where appropriate, noncognitive as well as cognitive functions. Empirical evidence relevant to the analyses should include results of investigations of the degree to which test scores are related or unrelated to other variables in ways implied by intended interpretations.

4. Criterion-related validation should be used only when technically sound and relevant criteria are available or can be developed and when other conditions affecting feasibility warrant the study.
   a. Criterion-related validation should involve as many performance variables as necessary to permit evaluation of the effectiveness of test scores for predicting the types of behavior they are intended to measure.

   b. Criterion-related validation should not combine variables to form a single criterion measure unless such a procedure is justified by logical considerations or empirical evidence or the practical requirements of the intended use of the results.

   c. Criterion data should be collected in a way that permits an assessment of the reliability of each criterion variable, but with the understanding that there may be several sources of irrelevant variation (e.g., sampling of criterion content, source of criterion ratings or data, and so forth).

5. Interpretations of correlations between test scores and criterion variables should take into account such factors as sample size, criterion reliability, possible restriction in the range of scores obtained in the validity study sample, and other contextual factors.

6. The method(s) by which any validation is accomplished should be fully documented; such documentation should include appropriate details such as the nature and reliability of the criteria, a description of the subjects used, the materials surveyed and the qualifications of the experts who made judgments regarding the appropriateness and importance of test content.

7. Where adequate methods are employed to insure equivalence of scores on alternate forms, it is not necessary that each new form be validated. New validation studies should be made if revised tests have substantial changes, such as different item types, or if they sample a revised performance domain.

8. When appropriate and feasible, the validity of a test should be investigated separately for subsamples of the test-taking population.

9. When a name of a test is established, it should not imply more than is justified by evidence of validity.

10. Information should be made available to institutional and agency users that would be of assistance to them in planning and conducting local validity studies.

TEST USE

Principle

Proper and fair use of ETS tests is essential to the social utility and professional acceptance of ETS work.

Policies

A. ETS will set forth clearly to sponsors, institutional or agency users, and examinees the principles of proper use of tests and interpretation of test results.

b. ETS will establish procedures by which fair and appropriate test use can be promoted and misuse can be discouraged or eliminated.

Procedural Guidelines

1. Program publications should:
   a. describe appropriate uses and caution against potential misuses of program tests;
   b. explain clearly that test scores reflect past opportunity to learn and discourage test interpretations that go beyond reasonable inferences from test performance;
   c. emphasize that an individual's test score should be interpreted in the context of other information about him or her;
   d. provide appropriate information about test content, difficulty, and purpose to help the institutional or agency user select instruments that meet the measurement requirements of the situation and avoid selecting, requiring or using inappropriate tests;
   e. invite institutional or agency users to consult with the program sponsor and/or ETS about their current or intended uses of ETS-developed tests and identify the offices to be contacted for this purpose;
   f. summarize results of research relevant to the use of the test or cite references in which such results are reported;
   g. describe adequately and clearly scale properties that affect score interpretation and use;
   h. advise institutional or agency users that decisions about the application of single or multiple prediction equations, based on distinguishing characteristics such as sex, ethnic group or curricular emphasis or training, should be preceded by careful examination of social, educational and psychometric factors;

i. advise institutional or agency users that if examinee grouping based on test scores is practiced, provision should be made for frequent review of group assignments to determine actual performance;

j. stress that pass-fail or cut-off scores established for such purposes as admission, credit, or certification should be used as a basis for decision-making only if the institutional or agency user has a carefully developed rationale, justification, or explanation of the cutting score that is adopted; and

k. encourage institutional or agency users to reexamine cut-off score policies periodically to minimize or eliminate possible disproportionate exclusion of members of any group such as men and women drawn from diverse backgrounds (e.g., major ethnic, handicapped and other subgroups of the population of interest) in the face of other evidence that would predict their success or indicate their competence.

2. Special (nonprogram) publications should be developed and disseminated by ETS to promote fair use of tests and discourage misuse of tests.

3. Complaints or information about questionable interpretation or use of reported scores should be investigated by means of procedures designed for detecting misuse. Such procedures should be documented, and records should be kept of such complaints and their disposition.

4. In cases where a clear misuse is brought to its attention, ETS should inform the sponsor and the institutional or agency user of ETS's opinion as to the misuse and seek voluntary correction of the misuse. If reasonable efforts to seek voluntary correction are not successful, ETS, in conjunction with the sponsor, should take steps to determine whether to continue supplying tests or reporting scores to the institutional or agency user.

TECHNICAL ASSISTANCE ADVICE, AND INSTRUCTION

Principle

ETS is dedicated not only to providing measurement programs and conducting research, but also to promoting increased understanding of measurement and test use.

Policies

A. ETS will develop and offer instructional programs in the areas of measurement, evaluation, and related research through such forms as publications, seminars, in-service training, intensive residence courses, workshops, internships and conferences. ETS may undertake these activities independently or in cooperation with other agencies, professional groups or educational institutions.

B. ETS will provide advice and information on measurement-related issues and about ETS programs, research and services. In this activity, ETS will work, where feasible, in collaboration with other professional organizations that show a concern about measurement.

C. ETS will respond promptly to requests for advice, instruction and technical assistance related both to programs and services offered by ETS and to the related areas of educational measurement, evaluation and research.

D. ETS will conform to high standards of accuracy and professionalism in its advisory, instructional and technical assistance activities.

E. ETS will provide advice, instruction and technical assistance to clients from the private and public sectors and from foreign and domestic government agencies to the extent that such services are consistent with ETS areas of expertise, meet accepted professional and ethical standards, and reflect an understanding of and respect for cultural differences.

F. ETS will endeavor to promote increased understanding of the purposes and procedures of testing among professional groups and in the public sector; ETS will make this effort both independently and in cooperation with other organizations that share this responsibility.

Procedural Guidelines

1. ETS's offices should offer advice, instruction and technical assistance, the staffing for such services should be determined by the nature of the services and the expertise required.

2. The special requirements of audiences with varying needs, interests, cultural backgrounds and levels of knowledge should be considered when ETS provides technical assistance, advice, or instruction.

3. New developments in research or testing should be considered when technical assistance, advice and instruction are offered.

4. Technical assistance, advice and instruction offered to institutions or agencies should include guidance on how to use other information about examinees (such as previous academic performance, English as a second language, and family or cultural background factors) in conjunction with test scores.

5. Comprehensive collections of reference materials relating to tests, measurement, evaluation and related research should be developed, maintained and made available to all ETS staff members and, when appropriate, to professional groups and individuals outside the organization.

GLOSSARY OF TERMS

Accuracy. The extent to which a principal product conforms to its specifications or correctly reflects the source data within the specified limits of reliability.

Client. (See Sponsor)

Consent. Permission granted by an individual or that individual's parent or guardian to the use or release of data held by ETS; such permission granted upon receipt of a reasonable explanation of the purpose of the use or release and a reasonable explanation of the manner in which the results will be reported.

Critical Information. Information that will be used to draw important inferences (a) about the sponsor, ETS-appointed external committees, institutional or agency user, examinee, subject or respondent, or (b) by the sponsor, institutional or agency user, examinee, subject or respondent and which, if incorrect, could be harmful.

Distributive Scale. A scale that is defined to yield either a specified score distribution or a specified mean and standard deviation for a particular group of examinees.

ETS Board of Trustees. The ETS Board of Trustees is the governing body of ETS. There are 17 trustees. Sixteen are elected for four-year terms. New members of the Board are elected by current trustees. The President of ETS is an *ex officio* member.

ETS-held Program Data Files. Information about individuals and institutions held by ETS and derived from ETS-provided services of collection, processing, storage, retrieval and dissemination.

ETS-held Research Files. Information held by ETS and generated through ETS-conducted research intended to result in the development of new or improved techniques and materials for application in such areas as classroom instruction, evaluation of progress toward educational goals, counseling of students, and decision-making of school administrators.

Examinee: An individual who takes a test, developed and or administered by ETS.

Institutional or Agency User: An organizational recipient of ETS-processed or produced information.

Intermediate Product: Materials that are not released externally, but that are necessary to the production of the principal product.

Nondistributive Scale: A scale that is defined without reference to the observed test performance of a particular group.

Nonnormative Scale: A scale that is based on the performance of any conveniently available subgroups of examinees for whom the test is appropriate. A score on a nonnormative scale is not intended to convey information about an examinee's standing in relation to a defined population.

Normative Scale: A scale that is based on the test performance of a sample of examinees, selected as prescribed by a specified design, from a clearly defined population. A score on a normative scale is intended to convey useful information about the performance of a particular examinee in relation to the performance of that population.

Principal Product: ETS-produced or processed materials (e.g., annual reports, performance data, score reports and admissions tickets) that are released or transmitted to a sponsor, ETS-appointed external committee, institutional or agency user, examinee, subject or respondent, pursuant to a contract or published commitment. Standards with respect to accuracy and timeliness are applicable to principal products.

Principles for the Validation and Use of Personnel Selection Procedures. Division of Industrial-Organizational Psychology, American Psychological Association. Dayton, Ohio: The Industrial-Organizational Psychologist, 1975.

Respondent. An individual who provides data to a research project in a manner and for a purpose different from either examinees or subjects.

Sponsor. Educational, professional or occupational associations, federal, state or local agencies, public or private foundations which contract with ETS for its services. This category includes their governing boards, membership, and appointed committees or staff.

Standards for Educational and Psychological Tests. American Psychological Association (APA), American Educational Research Association, and National Council on Measurement in Education. Washington, D.C.: APA, 1974.

Subgroup. A part of the larger population which is definable according to various criteria as appropriate, (e.g., by sex, race or ethnic origin, training or formal preparation, geographic location, income level, handicap and/or age).

Subject. An individual who participates in an ETS laboratory or experimental research project.

Testing Program. A set of arrangements under which examinees are scheduled to take a test under standardized conditions, the tests are supplied with instructions for giving and taking them, and arrangements are made for scoring the tests, reporting the scores, and providing interpretative information as part of a comprehensive ongoing service. A program is characterized by its continuing character and by the inclusiveness of the services provided.

Timeliness. The degree to which a principal product is released or delivered to its recipient within a predefined schedule.

ETS STANDARDS FOR QUALITY AND FAIRNESS

0151302 • Y101P3 • 292212 • Printed in U.S.A.

Appendix D

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

October 30, 1981

SUMMARY REPORT ON TEST DISCLOSURE REQUESTS, TEST ITEM CHALLENGES, HAND-SCORING REQUESTS, AND TEST FEES

Educational Testing Service and its client boards voluntarily have implemented nationwide the provision for test disclosure embodied in New York state law. The summary which follows documents the scope of this voluntary action for the 21-month period from January 1, 1980 to September 30, 1981. It is tangible evidence of what has been accomplished for openness in testing without federal legislation.

Test Disclosure Requests (January 1, 1980 to September 30, 1981)

Tests Disclosed:

Scholastic Aptitude Test, Preliminary Scholastic Aptitude Test,

Graduate Management Admission Test, Graduate Record Aptitude Test,

Test of English as a Foreign Language

Number of Major Administrations:	48
Number of Major Administrations Disclosed:	29
Number of Test-Takers at Disclosed Administrations (excluding PSAT):	1,175,000
Number of Test Disclosure Requests:	48,743
Number of PSAT/NMSQT Test-Takers Receiving Automatic Disclosure:	1,292,700
Average Response Time to Disclosure Requests:	Less than 3 weeks

Note: SAT disclosed in New York only in this period; will be available nationwide beginning in 1981-82 testing year.

Law School Admissions Test not included because ETS does not handle disclosure requests for this test.

Test Item Challenges (January 1, 1980 to September 30, 1981)
(Includes Law School Admission Test and special administrations disclosed)

Number of Test Items Disclosed:	6,480
Number of Test Forms Disclosed:	40
Number of Test Items Challenged:	74
Number of Flawed Test Items Resulting from Test-taker Challenges:	3
(1 each for the LSAT, SAT, and PSAT/NMSQT)	
Number of Flawed Test Items Identified and Removed Before Scoring by Normal Review Procedures:	11
(GMAT-1, GRE-3, LSAT-1, SAT-4, TOEFL-2; 10 different test forms)	
Number of Tests Rescored:	938,000
Number of Scores Revised:	289,000

Hand Scoring Requests (July 1, 1980 to September 30, 1981)
(Available nationwide for all administrations)

Number of Candidates All Administrations (including SAT, GMAT, GRE, and TOEFL)	2,424,000
Number of Hand Score Requests:	1,241
Average Response Time to Hand Score Requests:	Less than 3 weeks

Test Fees for 1981-82

Scholastic Aptitude Test	$10.50
Graduate Record Examination	$24.00
Graduate Management Admission Test	$27.00
Test of English as a Foreign Language	$19.00
Law School Admission Test (Test and Registration)	$38.00

Appendix E

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08541

BOARD OF TRUSTEES, 1980-81

Virginia B. Smith, Chairman
President
Vassar College
Poughkeepsie, NY

Paul F. Sharp, Vice Chairman
President Emeritus and Regents' Professor
Center for Studies on Higher Education
The University of Oklahoma
Norman, OK

Gregory R. Anrig
President
Educational Testing Service
Princeton, NJ

Charles R. Bowen
Director
Plans and Program Administration
in University Relations
International Business Machines
Corporation
Armonk, NY

Ernest L. Boyer
President
The Carnegie Foundation
for the Advancement of Teaching
Washington, DC

John T. Caldwell
Chancellor Emeritus
North Carolina State University
Raleigh, NC

Alfredo G. de los Santos, Jr.
Vice Chancellor for Educational
Development
Maricopa County Community
College District
Phoenix, AZ

Norman C. Francis
President
Xavier University of Louisiana
New Orleans, LA

Dorothy E. Gregg
Corporate Vice President
Communications
Celanese Corporation
New York, NY

F. Sheldon Hackney
President
University of Pennsylvania
Philadelphia, PA

Kurt M. Hertzfeld
Secretary to the Investment Committee
Amherst College
Amherst, MA

Elizabeth D. Koontz
Assistant State Superintendent
for Teacher Education
State Department of Public Instruction
Raleigh, NC

Charles D. O'Connell, Jr.
Vice President and Dean of Students
The University of Chicago
Chicago, IL

A. Kenneth Pye
Chancellor
Duke University
Durham, NC

Tomas Rivera
Chancellor
University of California
Riverside, CA

Stephen H. Spurr
Professor
Lyndon B. Johnson School of Public Affairs
The University of Texas
Austin, TX

Donald M. Stewart
President
Spelman College
Atlanta, GA

October 1981

Mr. WEISS. I have a few questions I would like to ask you, and at the outset I would ask also that in the event additional questions occur to members of the two subcommittees, many of whom are not here today because of other business, they may be submitted to you in writing after the close of this particular set of hearings and you will have 10 legislative days in which to respond.

Mr. ANRIG. I'd be happy to do so, Mr. Chairman.

Mr. WEISS. Fine. You had indicated that you've been working with the three major organizations that are involved in setting standards in your area—the American Psychological Association, the American Research Association, National Council on Measurement and Education. Is that a change in policy? Has that not been the case up to this point?

Mr. ANRIG. Mr. Chairman, there are two parts to your question. The first is that the existing standards of those three organizations independently arrived at by those three organizations have been incorporated into the ETS standards for quality and fairness that I referred to in my testimony. Those standards, however, were developed a number of years ago by those three associations. By their own decision, quite separate and apart from any organization in the testing field, they have decided to update and revise those standards. That updating and revision is going on now independent of any testing organization. It is being sponsored and directed by the three associations. What my testimony says is that our organization, Educational Testing Service, is willing to commit ourselves to those standards because we would do no less whether or not we felt the pressure to do so. We are prepared to do so when those standards come out. They are currently being developed independent of any testing organization.

Mr. WEISS. Right, had there been standards previously?

Mr. ANRIG. Yes, sir.

Mr. WEISS. And had ETS adhered to those standards previously?

Mr. ANRIG. That is correct. The standards of those three associations that have been currently in effect were used as the basis for developing what I referred to as the ETS standards for quality and fairness, and that's what we tested out over the last couple of years. We feel that we have met those standards and indeed exceeded that and are committing ourselves to go that far by corporate policy and beyond.

Mr. WEISS. I have my hands full taking on the testing agencies rather than starting another round with them, but if their standards have been ones that you've been adhering to I sort of question how diligent they have been in setting standards.

Mr. ANRIG. Mr. Chairman, I think first of all when I say that ETS has been adhering to those standards, that doesn't mean necessarily that all other organizations are. I'm speaking now again only for Educational Testing Service. Those standards, if you have your staff take a look at them or to meet with any members of those three associations, are in my judgment profoundly important and strong standards for quality and accuracy of testing. They don't cover all the areas with which you have addressed yourself, but they do cover many of the areas that underly your concern about testing. And I think they are by all judgments of the professionals in the area very high standards.

Mr. **Weiss**. Okay, have you conveyed the proposal and challenge that is incorporated on page 8 of your testimony to any of the other testing organizations, or is this the first public setting forth of that proposal and challenge.

Mr. **Anrig**. Mr. Chairman, I didn't want to take away in any way from the prerogatives of your committee or hearing. I wanted to express that first here, but I did have delivered to some of the other organizations that are affected by this, a copy of my testimony yesterday afternoon. So, they have been informed about it, but the formal announcement of it I wanted to reserve for today's hearing, out of respect to you.

Mr. **Weiss**. It's too soon yet to know what kind of response you will be getting from the other organizations.

Mr. **Anrig**. I think it will be a wonderful question for you to ask of them.

Mr. **Weiss**. Well, they haven't even had a chance to think about it yet. In the course of the past couple of years, we've had a number of objections raised to what we were attempting to do and you touched on some of those. I wonder if we could sort of go through them again. The question of costs, for example. Have you had occasion yet to undertake a study of what the total cost is of question development so far as ETS is concerned, and how that correlates to the proposed increases in the costs to applicants to take that test?

Mr. **Anrig**. First, with regard to the correlation of the cost to the applicant fees, as you know, Mr. Chairman, there is required in New York State an audit as I understand it of the increase in fees in connection with the Levale bill. If it's not required at least it's being followed out. So there is now an audit being conducted of the college board expenses, specifically with regard to the SAT, and the results of that audit will be made available as soon as that's completed. That's being done by the college board independent of us.

The question is, are there costs connected with test disclosure and the answer to that clearly is, yes. Indeed, logic would indicate that there would be. If you disclose forms of tests you have to develop new forms and, indeed, new forms are being developed. For instance, in the graduate management admissions test prior to disclosure we had to have two forms developed annually. Now we develop four. With the graduate record examination two forms were developed annually prior to disclosure. Now we develop four. With the scholastic aptitude test, before disclosure we developed 10 forms each year, I'm sorry, 7 forms each year, now we're developing 10. That means that just at ETS alone the number of people who work on test development logically has to be expanded to do that. From a number of 42 to 63 just during this period of time. A 50-percent increase to comply with the law.

Mr. **Weiss**. A 50-percent increase in what?

Mr. **Anrig**. In terms of the number of staff to develop the tests.

Mr. **Weiss**. Yes, but that still leaves open the question as to what percentage of the total amount of moneys that are taken in by ETS are in fact spent on test development and what the percentage of increase is due to the additional work that's required of ETS.

Mr. **Anrig**. OK, those are two different questions. First, with regard to every dollar that is spent by a test taker for the four major programs that we administer—the college board admissions testing program, the graduate management admissions test, the graduate record examinations, and the national teacher exam—out of every dollar paid, 9 cents of that dollar is for test development. That doesn't really tell you much. That doesn't tell you what the cost of test disclosure is. That just tells you what portions of the overall costs of ETS go for test development. The actual cost of test disclosure as we've been able to estimate them, and I would have to say that this is an estimate, Mr. Chairman, because I anticipated you to ask that question and I asked that it be drawn together. We've hustled around, I've only been there for 2 months, so we had to go around and do this; but let me give you our best estimates of that, and these are good faith figures that are subject to change but I think they're generally correct.

In the first year, taking all the programs that we administer that were affected by test disclosure. Taking the internal costs that we were incurring in terms of responding to that, it cost me something to come here today for instance, you take all of the costs that were involved directly with test disclosure, and in 1979-80 that came to an estimate of a little over $1.2 million for ETS. In 1980-81 because then we were developing the newer forms, we had a total cost estimate of about $2.3 million. And for 1981-82, this year, because we're moving ahead in some of the programs to further test disclosure, we estimate our costs will be about $2.6 million. Now, that takes all of the costs that we could legitimately identify and we weren't stretching this, that is, I said, don't put anything that's on the edges of this thing. I want it to be, as much as possible, an accurate picture of our direct, actual costs of test disclosure. They are those figures that I've just spoken to you. Now, that may just be a small part of a slice of a pie that says 9 percent, but it's a lot of money. So, it's important for us to understand that the worthy purposes that you're trying to accomplish do cost money.

Mr. WEISS. I appreciate that. Now, put that into context. Tell me what the fees were prior to 1979 and tell me what the fees are currently.

Mr. ANRIG. Sure. In 1979-80 the fee for the Scholastic Aptitude Test was $8.25. In 1980-81 it was $9.25. In 1981-82, this year, it will be $10.50.

Now, I think I want to make a point on that, Mr. Chairman. Back, as we used to say where I lived in your district, as a matter of fact, we can get it for you wholesale, and it is. If you take the costs in 1954-55, way before test disclosure, it cost $6 to take the admissions testing program. If you just put inflation on that, nothing else, just inflation from 1954-55, the fee this year should be $17.50. It is $8.25, or at least in 1979-80 it was $8.25.

Mr. WEISS. Commissioner, don't take me back to 1954. Our efforts to impose additional requirements on you only go back to 1979.

Mr. ANRIG. Yes, I gave you those figures.

Mr. WEISS. Right. If I'm correct, what you're saying is it's gone up from $8.25 to $10.50, right?

Mr. ANRIG. That's correct, for the scholastic aptitude test.

Mr. WEISS. Right. So there's an increase of somewhere around 25 percent.

Mr. ANRIG. Over a 3-year period, during the time when the inflation rate was going at a fairly high rate also.

Mr. WEISS. But the actual cost increase as a result of the additional burden was perhaps 9 percent. We have figures indicating it may have been 5 percent.

Mr. ANRIG. No, that latter part is not correct, Mr. Chairman.

Mr. WEISS. Tell me where it's incorrect.

Mr. ANRIG. The 9 percent that you're using is in reference, I believe, to the point I said that for every dollar spent on tests, 9 cents of that dollar is for test development.

Mr. WEISS. OK, tell me what the percentage of increase has been.

Mr. ANRIG. I don't know, I majored in history.

Mr. WEISS. Would you do the committee a favor? Find somebody that majored in math.

Mr. ANRIG. And did well on the SAT.

Mr. WEISS. And develop a presentation in response to the question that I asked. You know the overall context. Don't limit yourself to whatever flaws there may have been in my question, but take the overall cost-increased expenditure area and try to develop a response so that, in fact, we have a full statement directed as to what the additional increases in fees have been, what the additional costs have been, with the appropriate breakdowns and so on, OK?

Mr. ANRIG. I'll be happy to do that.

[Answers to Mr. Weiss' questions follow:]

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08541

Gregory R. Anrig
President

December 10, 1981

The Honorable Carl Perkins, Chairman
Subcommittee on Elementary, Secondary,
and Vocational Education
The Honorable Paul Simon, Chairman
Subcommittee on Postsecondary Education
320 Cannon House Office Building
Washington, D.C. 20515

Dear Messrs. Perkins and Simon:

During the hearing on H.R. 1662 on November 4, Congressman Weiss requested additional information from ETS on several points: The number of students tested by ETS in major admissions programs, appeal procedures which students can use in test security cases, and costs related to disclosure. I am providing that information below.

1. Number of students tested. The attached table (Table 1) provides candidate volumes for major admissions and placement programs. The numbers represent candidates in each program. However, it should be noted that many students who take the Advanced Placement Exam also take the SAT and Achievement tests. Similarly, some students take more than one professional school exam, e.g., GRE and GMAT. Thus, the total of 4,119,400 is not an unduplicated one.

   The enclosed brochure, entitled "ETS in Fact," provides information for the 1979-80 testing year as well as additional general information about ETS programs.

2. Score validity procedures. We make every effort to give every student an equal chance during the administration of the test in the test center. We want to be sure some candidates do not have an unfair advantage over others. Thus, we take the matter of test security very seriously. At the same time, students whose scores are questioned through our security procedures must be assured that they are treated fairly.

   The procedures we use in investigating the validity of questioned scores are explained in the enclosed booklet entitled "Score Validity Procedures." In addition, some test sponsors provide for third party mediation should normal procedures fail to resolve the matter. As an

2. (cont.)

example, I am enclosing the policy statement adopted by ETS and the College Board which applies to tests sponsored by the College Board.

3. Fee increases related to disclosure.

It is important to understand that test fees are set by the policy boards and councils that sponsor each of the testing programs, not by ETS. It is the policy of each of the councils to keep fees as reasonable as possible but fee increases required vary from program to program depending on such factors as a) changes in program services, b) the number of people who take the test, c) inflation, and d) financial circumstances and reserve requirements of the sponsoring agency. Costs of disclosure are but one factor. Since the passage of the New York State law in July 1979, the cumulative fee increases to date in the programs administered by ETS which are covered by the law range from less than 12 percent to 116 percent.

Similarly, the portion of these increases that is directly attributable to disclosure varies among programs, depending on the number of test forms disclosed, number of test-takers, and changes in test development, equating and administration procedures. For example, for the Test of English as a Foreign Language, no change was required in the number of test editions developed annually or in the format of the test, whereas the Graduate Management Admission Test had a 100 percent increase in new test editions developed annually along with the development and implementation of a new, much more complex and expensive method of providing comparability of scores earned at different administrations. Although it is extremely difficult to separate the elements contributing to fee increases, our best estimate is that the increases in test fees in 1980-81 directly attributable to disclosure ranged from very little to more than 37 percent. Test fee increases for the current testing year range from 0 to 20 percent. The increases were due mostly to a combination of inflation and disclosure, with the larger percentage increases occurring in those programs which had to incorporate sharply increased test development cost and changed test formats. A table outlining test fees for the three most recent years is attached (Table 2).

4. Implications of H.R. 1662.

Implications of H.R. 1662 are not merely a matter of costs. Two sections, Sections 3 & 5, include requirements that not

4. (cont.)

only would increase costs but would present serious problems of feasibility at any cost.

Section 3

Although all programs administered by ETS already provide most of the information called for in Section 3,[1] the requirement in Section 3(A)(3) for reporting to test subjects "the correlation between test scores and success in the career for which admission is sought" calls for information not now available. In the unlikely event that consensus could be reached on a definition of success, it would take many years to collect such information and at great expense Most test materials include statements saying that the tests are not designed to predict success in a given field and the career paths taken by individuals after completion of graduate or professional training are many and diverse.

In Section 3(4)(B), "a comparison of the average score and percentiles by major income groups" is called for. Some programs do not collect income information from candidates. Reporting such information would require revision of the application form, of the data base and the computer system for the program. For a test such as TOEFL (Test of English as a Foreign Language) which has many foreign candidates from many different countries, the information would be uninterpretable. For graduate and professional programs, some applicants are dependent on parental support while others are self-supporting with typically low incomes; we would be dealing with "apples and oranges."

Section 3(4)(C) calls for "the extent to which test preparation courses improve test subjects' scores on average." Meaningful statistical averages or percentages on the effects of special preparation are not available for most programs. TOEFL, for example, is an achievement test for which additional preparation would be expected to produce significant gains. The design for a research project and the collection of data for one year would cost about $75,000.

Assuming that the additional information called for could be provided, candidate informational bulletins

---

[1] See enclosed statement, "Side-by-Side Comparison of H.R. 1662, Provisions and Current Practices of Guidance and Admissions Tests Administered by ETS."

4. (Cont.)

would have to be expanded, with increased costs for printing and distribution. For TOEFL, it is estimated that the additional cost of distributing an expanded handbook worldwide alone would be about $92,000.

Section 5

The costs of implementing the full disclosure called for in this section would be twofold: 1) the cost to candidates of decreased flexibility, convenience and opportunities to take tests; and 2) dollar costs, resulting in higher test fees.

Because of the problems associated with disclosing tests for every administration in programs testing 5000 candidates or more, and technical difficulties associated with developing new tests for the small numbers of candidates served at special administrations such as those for Sabbath observers and the handicapped, it is highly probable that the policy boards would decide to eliminate some administrations and even drop some tests entirely, particularly from the College Board Achievement Tests and Graduate Record Advanced Tests.

We tried to estimate some test development costs using the assumptions that the current level of test administrations would be maintained and that no changes in length or format of tests would be required:

a. Admissions Testing Program of the College Board

1. SATs.

Annually 22 forms are used.
Seven SAT forms are disclosed.
Currently 10 new forms are developed each year.
Twelve additional ones would be needed, if administration for special populations such as handicapped are limited to one or two per year.
Additional cost of SAT test development: $1,320,000.
Additional cost per candidate, using 1981 volumes: 83¢.

2. Achievement Tests.

Ten Achievement tests would probably fall under provision of H.R. 1662. Each is administered 10 times per year.
On the average, one new form of each is produced per year.

4. (Cont.)

Nine new forms of each would be needed.
Additional cost of test development: $2,970,000.
Additional cost per candidate: $10.31.

3. Total additional cost of test development: $4,290,000.

b. Graduate Record Examinations

1. Aptitude Tests.

Annually 10 forms are used in major administrations.
There are more than 100 special administrations in large urban areas, the cost of which are not estimated here.
Three forms are disclosed.
Currently 3 new forms are developed each year.
Seven additional forms for major administrations would be needed (eliminating special administrations).
Additional test development costs: $1,155,000.
Additional cost per candidate: $4.40.

2. Advanced Tests.

Six tests would probably be covered by H.R. 1662.
Each is administered 8 times per year in regular administrations, as well as in numerous special administrations.
On the average, one new form of each is produced per year.
Seven additional forms of each would be needed for regular administrations.
Additional test development costs: $2,226,000.
Additional costs per candidate (estimated volume 1980-81): $37.10.

3. Total additional test development costs: $3,381,000.

c. Graduate Management Admission Tests

All 4 regular administrations are disclosed currently.
Three Monday administrations plus those for handicapped students are not disclosed.
Currently 4 new forms are developed per year.
Four additional new forms would be needed, assuming handicapped testing would be limited to once per year.

4. (Cont.)

Additional test development costs: $600,000.
Additional costs per candidate served on Monday and in handicapped administration: $545.
Additional costs spread over all candidates: $2.73.

The assumptions of no changes in frequency of administrations and in format or length of tests are unrealistic. For example, the disclosure of all SAT forms would require changes in the PSAT/NMSQT program. At the present time, two new forms of the PSAT/NMSQT are developed each year from two undisclosed SATs. If undisclosed SATs are no longer available, new questions must be developed and pretested specifically for PSAT/NMSQT and new methods for maintaining score comparability introduced at a presently unknown cost and at least a two-year lead time.

Currently, the voluntary national policy of disclosure provides the flexibility required to maintain services to special populations, e.g., handicapped, Sabbath observers, military personnel, students needing make-up administrations. Programs such as GRE and SAT now use tests equated on large administrations for those special populations. Present methods of making scores comparable (as well as one being developed) require a much larger number of test-takers than are available in the non-Saturday administrations. Other new methods would require considerable research and, if effective, the lengthening of the tests.

The problems for the Achievement and Advanced tests include those mentioned above plus the formidable task of producing a sufficient quantity of quality questions on the same topics in a subject area, should 8 to 10 times the number of new forms have to be produced each year.

Consequently, should H.R. 1662 be passed in its current form, some of the tests may be dropped from the list of offerings and administrations of others limited to one or two per year. Moreover, for New York there is an ironic twist. Should the policy boards decide to eliminate any small volume administrations for Sabbath observers, there would be a concurrent elimination of the companion, regular Saturday administrations in New York, because an amendment to the New York law requires that an equal number of opportunities to take the test must be available to Sabbath observers.

As you can see, the implications are far reaching. While the concept of disclosure of test content seems straightforward and one we support--the variations among programs, services to students, and arrangements for special situations are such that sweeping legislation, however straightforward in intent, will have unintended effects. I repeat my strong belief, as stated in my testimony, that the goal of openness in testing can be and is being achieved without further legislation.

I hope this information responds adequately to the questions raised. Please let me know if I can be of further assistance.

Sincerely yours,

Gregory R. Anrig
President

GRA:jd

TABLE 1

ETS TEST CANDIDATE VOLUMES FOR SELECTED NATIONAL PROGRAMS

Testing Program	Sponsor	Candidate Volume/ 1980-81
ELEMENTARY & SECONDARY LEVEL		
Secondary School Admission Test (SSAT)	Secondary School Admission Test Board, Inc.	52,500
COLLEGE LEVEL		
Admissions Testing Program (ATP)	College Board	1,556,100
Scholastic Aptitude Test only		1,282,100
Achievement Tests only		76,000
SAT and Achievement Tests		198,000
Advanced Placement Program (APP)	College Board	134,000
College-Level Examination Program (CLEP)	College Board	247,000
National Teacher Examinations (NTE)	National Teacher Examinations Policy Council	75,000
Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT)	College Board/National Merit Scholarship Corporation	1,295,000
Test of English as a Foreign Language (TOEFL)	College Board/Graduate Record Examinations Board	225,300
GRADUATE & PROFESSIONAL LEVEL		
Graduate Management Admission Test (GMAT)	Graduate Management Admission Council	219,500
Graduate Record Examinations (GRE)	Graduate Record Examinations Board	285,000
Law School Admission Test (LSAT)	Law School Admission Council	110,000
Total		4,119,400

POLICIES AND PROCEDURES FOR ARBITRATION OF TEST SECURITY ACTIONS FOR COLLEGE BOARD PROGRAMS

Test takers whose scores have been questioned deserve assurance that their rights to confidentiality and fair treatment are fully protected. Over the years, ETS and the College Board have tried to ensure that the interests of both test takers and score users are given balanced consideration.

Test scores are cancelled or withheld only if there is a reasonable basis for questioning them and only after a thorough review of the pertinent facts - including an opportunity for the test taker to provide additional information. Such actions are intended to ensure equity, not to be punitive; further, they are kept in strict confidence. The ETS Board of Review, a group of senior members of the professional staff, reviews all information about scores of questionable validity including information that the candidate or his or her designee provides. Only if there is unanimous agreement among three Board of Review members that there is a reasonable basis for questioning a score is the candidate notified of the review, given the preliminary decision, and informed of the options.

Candidates whose scores are questioned by ETS are offered the following options:

1. to provide additional information to the Board of Review that the candidate believes would lead to a different decision;
2. to retest to confirm the questioned scores;
3. to cancel the scores with no record maintained of any questions raised;
4. to submit all information on the questioned scores to the colleges that the candidate designates.

The procedures described above are based on established practice and procedures for dealing with test scores of questionable validity in order to ensure fair treatment and confidentiality. In establishing the following additional procedures, ETS and the College Board provide an option for candidates to go beyond the existing procedures by appealing an ETS Board of Review decision to an independent third party.

Arbitration Procedures

1. The candidate is informed by letter of the Board of Review's preliminary decision that there is a reasonable basis for questioning the validity of the candidate's score and is given the four options specified above. In the same letter the candidate is told of the availability of third-party arbitration if additional information provided by the candidate to the Board of Review does not persuade it to change its decision.

2. If the Board of Review concludes that doubt still exists about the authenticity of the score after the candidate submits additional information, the candidate is reminded that appeal to an independent third party is available and is told to contact ETS if he wishes to initiate this process.

3. On request to inititate the arbitration process, ETS will send a copy of an Arbitration Agreement Form (see attachment A) to be signed by the candidate.

4. Upon receipt of the signed Arbitration Agreement Form for third-party arbitration, ETS will forward the request to the American Arbitration Association with a summary of all information, along with copies of supporting documents, that led to the decision to question the validity of the candidate's scores.

5. Upon receipt from ETS, the American Arbitration Association will send to the candidate a duplicate set of all the materials provided by ETS that led to the decision to question the validity of the scores and will give the candidate twenty-one days to reply. The candidate will be informed that, while under normal circumstances, the costs of arbitration will be borne by the College Board, the arbitrator may at his discretion charge the candidate for a portion of the costs (not to exceed $200) if the arbitrator finds that the appeal is without reasonable foundation.

6. All information received by the American Arbitration Association from one party will be provided to the other party with the opportunity for a response.

7. Ordinarily the arbitrator's decision will be made within ten working days of receipt of the necessary information from the candidate. (If no response is received from the candidate within the allotted time, the arbitrator will make a decision based on the documents already submitted.) Ordinarily, the decision will be based on information submitted by mail. However, the arbitrator will consider a request for a personal hearing if there are compelling reasons to do so. The cost incurred by the candidate for such an appearance will be borne by the candidate. If a personal hearing occurs, both sides will be invited to appear. The results of the decision will be forwarded in writing to the candidate and ETS.

8. If the candidate's appeal is denied, the candidate's score will be cancelled. If the candidate's appeal is upheld, ETS will inform the candidate by letter of the specific actions taken to implement the arbitrator's decision, including that the original score reports have been released to all recipients originally designated by the candidate along with, if necessary, an appropriate letter of explanation.

Periodic Review

The policies and procedures described in this paper should be reviewed periodically and may be amended or revised based on experience and changing circumstances.

July 1981

ATTACHMENT A

Agreement for Submission to Arbitration of Test Security Actions

On ________(date)________, I took the ____(name of exam)____ examination administered by Educational Testing Service at ________________ test center.

ETS is withholding release or, if already released, planning to cancel my score(s) because of a decision by its Board of Review that there is a reasonable basis for questioning the validity of the score(s). I have been informed of my options, namely: (a) retest to confirm the questioned scores, (b) submission of additional information to the Board of Review that I believe would lead to a different decision, (c) authorizing the cancellation of my scores, with no record maintained of any questions raised, (d) release of all information on the questioned score to those colleges I have designated as score recipients. I have submitted additional information to the Board of Review which has not persuaded it to revise its original decision that there is a reasonable basis for questioning the validity of the score(s). I am now choosing to submit the matter to independent third-party review with the following conditions:

1. I understand that the third-party arbitrator will decide only whether ETS has a reasonable basis to question the validity of my scores and to withhold or cancel them.

2. I understand that the decision of the arbitrator is final, and I agree, as a condition of submission of the issue as mentioned above, to be bound by the decision.

3. Upon receipt of this agreement signed by me, ETS will forward this agreement, together with a summary of all information in its possession leading it to question the validity of my score(s), plus copies of supporting documents, to the American Arbitration Association (AAA).

4. The AAA will then write to me and give me the opportunity to submit in writing my position that the questioned scores are valid, as well as any additional relevant materials supporting my position. The AAA will also send me a copy of all materials submitted to them by ETS and will send a copy of materials submitted by me to ETS.

5. I will have twenty-one days from the date on the AAA's letter to me (as mentioned in Paragraph 4, above) to submit my information. Following receipt of my information, the arbitrator will reach a decision within ten business days. I will receive a copy of that decision, in writing, from the AAA.

6. If I do not submit the information requested by the AAA mentioned in Paragraph 5, above, within the allotted time, the arbitrator will make a decision based on the information available at the expiration of twenty-one days from the date of the AAA's letter to me offering me a chance to submit information.

7. I understand that the arbitrator's decision will ordinarily be based on information submitted by mail. However, the arbitrator will consider requests for a personal hearing if there are compelling reasons to do so. If a personal hearing occurs, both sides will be invited to appear. I understand that any costs incurred by me in such an appearance will be my responsibility.

8. If the arbitrator's decision is in my favor, ETS will inform me in writing of the specific actions being taken to implement the arbitrator's decision, including that the original score reports have been released to all recipients designated by me.

9. If the arbitrator's decision is not in my favor, ETS will cancel my score.

10. I understand that, under ordinary circumstances, the costs of the arbitration process will be borne by the College Board. I further understand, however, that, should the arbitrator find my appeal is without reasonable foundation, I may be charged for a portion of these costs not to exceed $200 at the discretion of the arbitrator.

I have read, understood, and agree to be bound by this agreement. I am aware of my right to have the advice of anyone I choose before I sign this agreement.

____
(Candidate)

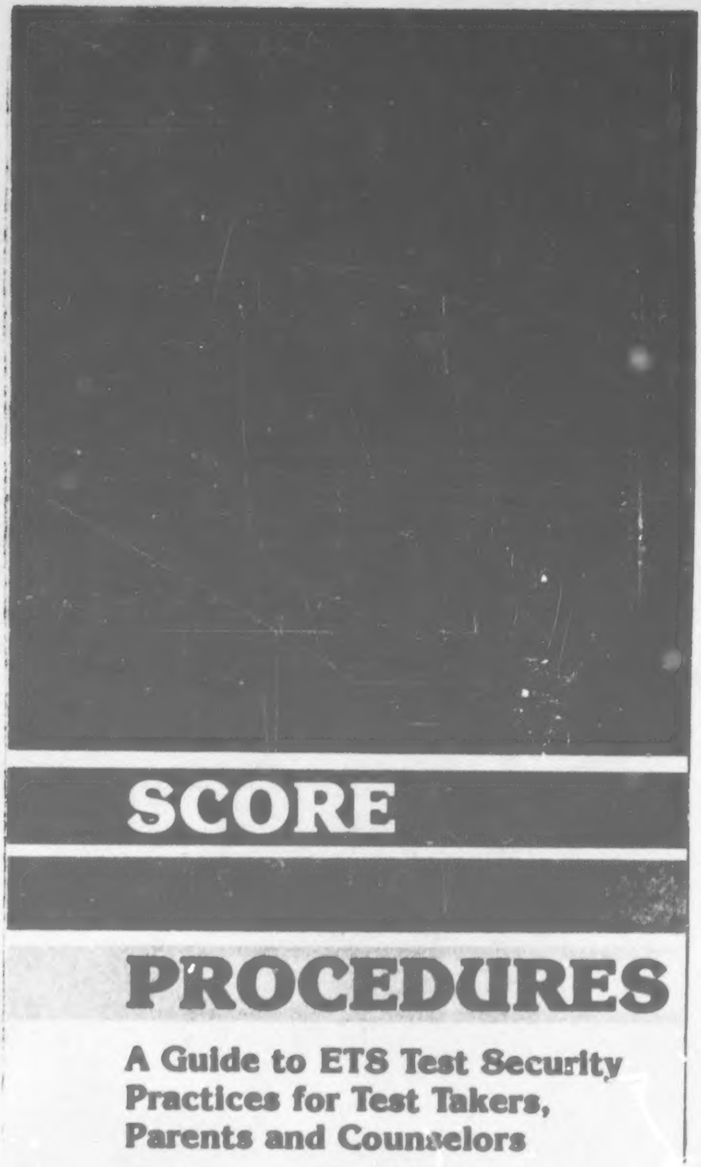
____
(Parent or guardian of the candidate,
if the candidate is not 18 years of age)

____
(Date)

TABLE 2

TEST FEES

	1979-80	1980-81	1981-82
SAT	8.25	9.25	10.50
New York Surcharge		1.75	.50
GRE	14.00	20.00	24.00
GMAT	12.50	23.50	27.00
TOEFL			
International Testing Program	17.00	19.00	19.00
Special Center Testing Program	25.00	27.00	27.00
LSAT AND LSDAS			
Basic Registration Fee	----	16.00	18.00
LSAT	15.00	15.00	20.00
LSDAS	18.00	15.00	15.00
LSAT only	15.00	31.00	38.00
LSDAS only	18.00	31.00	33.00
LSAT + LSDAS	33.00	46.00	53.00



SCORE PROCEDURES

A Guide to ETS Test Security Practices for Test Takers, Parents and Counselors

This pamphlet is intended to familiarize test takers—and their parents and other persons from whom they may seek advice—with the procedures for reviewing test scores of questionable validity. Examinees and other affected parties are entitled to know why certain scores have been identified for review and the procedures used by Educational Testing Service (ETS) and testing program sponsors either to verify or invalidate scores that have been questioned.

Educational Testing Service • Princeton, N.J. • 08541
ETS is an Equal Opportunity Employer.
Copyright © 1981 by Educational Testing Service.
All rights reserved.

A Matter of Equity

Test takers whose scores have been questioned deserve assurance that their rights to confidentiality and fair treatment are being fully protected. Over the years ETS has made a diligent effort to ensure that the interests of both test takers and score users are given balanced consideration.

Our concern for equity is precisely why we are stringent in matters of test security. Test takers who attempt to enhance their scores through improper means, such as impersonation or copying answers, are abusing the rights of the vast majority of those who take tests unassisted and in good faith. The first obligation of ETS and the test sponsors is to ensure that some candidates are not permitted to have an unfair advantage over any others. We also have an obligation to institutions that receive scores to provide them with information that accurately reflects the abilities of test takers. Thus, when persons register to take an ETS-administered test, they sign, in effect, a contract acknowledging that ETS or the test sponsor for whom we are an agent (such as the College Board or Graduate Record Examinations Board) have the sole right to cancel any test score if there is reason to question its validity.

Test scores are canceled or withheld only if there is a reasonable basis for questioning them and only after thorough review of the facts pertinent to each case—including information presented by the test taker. Such actions are not intended to be punitive; they are kept in strict confidence. Examinees who have had their scores questioned are offered a retest or, depending on the policies established by the test sponsor, an opportunity to appeal the decision

Q:A:

Note: These are general procedures for **most** ETS tests. The Law School Admission Council and certain other test sponsors have other but similar procedures. Affected test takers are advised when different procedures are used.

Following, in question and answer form, is an explanation of ETS test security procedures and steps that can be taken if test scores have been questioned:

Q: Who assembles information about scores of questionable validity?

A: Information is assembled by the ETS Test Security Office under policies established by the ETS Board of Review, a group of senior members of the professional staff responsible for dealing with such issues.

Q: How are questionable scores first identified?

A: Such scores are most commonly identified through a computer check of all score differences following each test administration. If the comparison of the most recent score with a previous score (or scores) reveals an unusual difference, available information is assembled and automatically reviewed for handwriting differences, indications of copying, an exchange of answer sheets, or for other signs of possible irregularities. Questionable scores also are brought to the attention of the Test Security Office by score recipients who cannot reconcile an apparent inconsistency between a test score and other information known about the candidate, such as previous test scores or academic performance. Other test takers or test center supervisors who suspect or have observed misconduct also report such information to the ETS Test Security Office. Occasionally ETS receives anonymous letters alleging irregularities. Such cases are reviewed, but the specific content of such letters is not presented to the Board of Review for its consideration.

Q: Is a large score difference, in itself, sufficient reason to cancel a score?

A: No. Such a score difference must be accompanied by at least one other indication of an irregularity.

Q: What happens to test scores while they are being reviewed?

A: Scores that have not been reported are withheld pending final outcome of the review, and are released only if cleared. If the scores have already been reported, score recipients are not advised by ETS of the score status until final determination of the case.

Q: What happens during a review?

A: If the Test Security Office concludes, upon review, that there is no reason to question the score, the inquiry goes no further and, if the score has not been released, it is reported normally. If the score has already been reported, any inquirers are informed that the score is valid. (Approximately two thirds of ETS reviews end at this stage of the process.)

Q: What if there is a reasonable basis for questioning the validity of a score?

A: If the Test Security Office finds that the information assembled indicates reason to question the validity of a score, the case is presented to the Board of Review with a recommendation for a specific course of action. The Board then examines the information and, if there is **unanimous** agreement among the quorum of three Board members that there is a reasonable basis for questioning the score, the examinee is notified of the Board's review and preliminary decision. At this time, examinees are also informed of the options open to them.

Q: And if there is not unanimous agreement by the Board?

A: If even one member of the Board determines that there is insufficient reason to question a score, it is cleared and reported according to normal procedures.

Q: What options are available to a test taker whose score has been questioned?

A: Three courses of action are open to the test taker:

1. He or she is entitled to take a retest under special supervision to confirm a questioned score.* If performance on the retest is reasonably close to the score in question, ETS will either affirm its validity to score recipients who may have questioned it (if the score has already been released), or report the score promptly to designated score recipients (if it has been withheld). A test taker who has been advised of possible discrepancies may request the score confirmation standards for his or her particular test from the Test Security Office. If the retest fails to confirm the questioned score, ETS will cancel the score and, if it had been previously released, so notify score recipients. The basic test fees, plus any additional fees, will be refunded.

   (A test taker who elects this option should be aware that he or she will be required to provide, in addition to regularly required identification, a separate and additional recent photograph which will be forwarded to ETS with the retest. The examinee also is required to name three responsible individuals who can identify him or her from this photograph. Each test taker may be asked also to provide a thumbprint as an additional means of positive identification at the time of the retest.)

---

*Because of administrative difficulties, persons tested outside the United States and Canada are not customarily offered a retest for the purpose of confirming a questioned score. They may, however, provide additional relevant information to the Board of Review for its consideration or authorize cancellation of the questioned score.

If an examinee elects this option he or she may request, prior to learning his or her retest results, that the retest score be retained as the score of record. There is no refund of fees when the retest score is reported in this way.

2. A test taker may authorize ETS to cancel the questioned score and remove it from the record. The basic test fee plus any additional fees will be refunded, and, if the score has already been reported, ETS will notify the score recipients that it has been canceled. Recipients will not be informed of the reasons for cancellation. If a score has been held pending investigation, no disclosure of the cancellation will be made to designated score recipients.

3. Test takers may provide the Board of Review with additional information for its consideration, such as other test results or academic records or a doctor's certificate attesting to physical impairments that might affect handwriting. (Character references are not generally considered by the Board to be relevant to the validity of a test score.) If the Board finds the information unconvincing, the examinee may submit additional information, take a retest, or authorize cancellation of the score.

A test taker who decides to take one of the options described should contact the Secretary of the Board of Review. If none of these options is chosen, the questioned score will be withheld and not reported by ETS. If the score has already been reported, score recipients will be notified only that the score has been canceled; no reference will be made to the reason for the cancellation.

Q: When must a test taker select one of these options?

A: If a score has been withheld, there is no deadline; the score in question is simply held in an inactive status until such time as the candidate decides to respond.

If a score has previously been released, the examinee is notified of a date by which he or she must respond. If there is no response by the stipulated date, the score will be cancelled, fees refunded, and score recipients so notified. Score recipients will not be informed of the reason for the cancellation.

Q: Is there anything else test takers should know?

A: The ETS Test Security Office encourages examinees to call for further information about the options described above. In addition, certain test programs have established procedures for third party mediation of test security cases. The Test Security Office can inform test takers of the specific test programs to which these procedures are applicable, and under what conditions.

Additional Options

ETS views the nature and results of test score validity investigations as strictly confidential. However, we recognize the possibility that, based on the available information, other parties could come to different conclusions about the validity of a test taker's score. Institutions are not bound by ETS decisions regarding questioned test scores, and may make their own judgments on the usefulness of a test score for their own decision making. Therefore, with a test taker's written authorization and the agreement of the institution(s) involved, all information pertinent to the investigation of the score will be made available by ETS to a representative of the designated institution(s).

While examinees should be aware that not all admissions officers may be willing to review the facts and make independent judgments, this is a course of action worth considering.

Finally, each test taker is encouraged to speak with his or her parents, counselor or other advisor before making a final decision regarding the options available to him or her. This is recommended if the test taker is a minor. The Test Security Office is ready to discuss any aspect of this matter with affected examinees, and the Board of Review will reconsider any case in light of such discussion if requested.

ETS IN FACT 1981

489

Educational Testing Service (ETS) is a nonprofit organization committed to development and administration of responsible testing and financial aid programs; creation of advisory and instructional services; and research on techniques and uses of measurement, human learning and behavior, and educational development and policy formation.

ACADEMIC TESTING PROGRAMS

ETS develops and/or administers tests for various sponsors in the United States and administers examination programs in some 156 other countries. Additionally, the organization supplies related services: scoring tests; recording, storing, and reporting test results; performing validity studies; undertaking program research. The numbers (to the nearest thousand) taking the tests in 1979-80 are listed for several major programs.

Testing Program	Sponsor*	Candidate Volume/1979-80
Elementary & Secondary Level		
Educational Records Bureau (ERB)	ERB	173,000
Secondary School Admission Test (SSAT)	SSAT Board	50,000
College Level		
Admissions Testing Program (ATP)	CB	1,527,000
		1,255,000

Test of Standard Written English (TSWE)		[**]
Achievement Tests only		76,000
SAT and Achievement Tests		196,000
Advanced Placement Program (APP)	CB	120,000
College-Level Examination Program (CLEP)	CB	266,000[***]
Comparative Guidance and Placement Program (CGP)	CB	[****]
National Teacher Examinations (NTE)	NTE Policy Council	76,000
Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT)	CB/NMSC	1,270,000
Test of English as a Foreign Language (TOEFL)	CB/GRE Board	282,000
Graduate & Professional Level		
Graduate Management Admission Test (GMAT)	GMAC	205,000
Graduate Record Examinations (GRE)	GRE Board	302,000
Law School Admission Test (LSAT)	LSAC	112,000

---

[*] CB—College Board. ERB—Educational Records Bureau, Inc.; ETS—Educational Testing Service. GMAC—Graduate Management Admission Council; GRE Board—Graduate Record Examinations Board. LSAC—Law School Admission Council; NMSC—National Merit Scholarship Corporation; SSAT Board—Secondary School Admission Test Board, Inc.

[**] TSWE—Administered in conjunction with the SAT

[***] Number of examinations administered

[****] CGP—Institutionally administered: 40,000; individually self-scored: 298,000

DIRECT LINES FOR TESTING PROGRAMS

ATP	■ 609-771-7600	**LSAT**	■ 215-968-1000	**SSAT**	■ 609-771-7530
GMAT	■ 609-772-7330	**NTE**	■ 609-771-7330	**TOEFL**	■ 609-771-7570
GRE	■ 609-771-7615				

Inquiries concerning other ETS testing programs should be directed to the Princeton Office, 609-921-9000. The cable address is EDUCTESTSVC; telex 84-3420.

FINANCIAL AID SERVICES

Financial Aid Service	Sponsor*	Volume/ 1979-80	For information, write:
Secondary Level School Scholarship Service (SSS)	NAIS	36,000	School Scholarship Service Box 599, Princeton, NJ 08541
College Level College Scholarship Service (CSS)	CB	2,850,000	College Scholarship Service Box 2700, Princeton, NJ 08541
Graduate & Professional Level Graduate & Professional School Financial Aid Service (GAPSFAS)	GAPFAC	88,000	Graduate & Professional School Financial Aid Service Box 2614, Princeton, NJ 08541

*CB—College Board; GAPFAC—Graduate and Professional Financial Aid Council; NAIS—National Association of Independent Schools

INSTITUTIONAL SERVICES

Institutional Service	Sponsor*
Secondary Level	
Questionnaire for Students, Teachers, and Administrators (QUESTA I and QUESTA II)	SSRP
College Level	
ATP Summary Reporting Service (SRS)	CB
Community College Goals Inventory (CCGI)	ETS
Institutional Functioning Inventory (IFI)	ETS
Institutional Goals Inventory (IGI)	ETS
Small College Goals Inventory (SCGI)	ETS
Student Instructional Report (SIR)	ETS
Student Reactions to College (SRC)	ETS
Student Search Service (SSS)	CB
Undergraduate Assessment Program (UAP)	UAP Council
Validity Study Service (VSS)	CB
Graduate Level	
Minority Graduate Student Locater Service (MGSLS)	GRE Board

*CB—College Board. ETS—Educational Testing Service; GRE Board—Graduate Record Examinations Board; SSRP—Secondary School Research Program, Inc.; UAP Council—Undergraduate Assessment Program Council.

OCCUPATIONAL AND PROFESSIONAL MEASUREMENT PROGRAMS

The ETS Center for Occupational and Professional Assessment (COPA) develops and/or administers written and performance tests for licensing, certification, continuing education/self-assessment, and selection in some 30 occupations and professions. These include:

- **Actuaries**
- **Architects**
- **Auto and truck mechanics**
- **Business forms designers**
- **City planners**
- **Construction code inspectors**
- **Contact lens dispensers**
- **Data processing personnel**
- **Dental auxiliaries**
- **Engineers**
- **Foreign Service Officers**
- **Public housing managers**
- **Insurance agents**
- **Lawyers**
- **Nurses**
- **Obstetricians and gynecologists**
- **Opticians**
- **Pharmacists**
- **Physicians**
- **Podiatrists**
- **Purchasing managers**
- **Radiologic technologists**
- **Real estate brokers and salespersons**
- **Respiratory therapists**
- **Social workers**
- **Travel agents**
- **U.S. Employment Service personnel**

OFFICE FOR MINORITY EDUCATION (OME)

OME coordinates ETS's continuing efforts to help minority students gain access to and benefit from education at all levels. OME collects information on barriers to minority achievement and identifies relevant research and services. OME provides consultation and training for educators and students in such areas as bilingual education, assessment, instruction, and test awareness.

TEST COLLECTION

The Test Collection is probably the largest open library of its kind in the world. It provides access to most tests and assessment devices published in the U.S., as well as many of those published in foreign countries. Instruments cited in the journal literature, but not available through publishers, are also collected, and some of these can be obtained in microfiche form from the Test Collection. In addition, information on test publishers is provided.

ERIC/TM CLEARINGHOUSE

Supported by the National Institute of Education, each of the 16 Educational Resources Information Center (ERIC) clearinghouses is responsible for collecting, cataloging, indexing, and abstracting unpublished research and development documents. Over 170,000 such documents on microfiche comprise the complete ERIC collection, which can be accessed through a computer retrieval system. The ERIC/TM Clearinghouse, located at ETS, is responsible for documents relating to testing, measurement, evaluation, learning theory, and human development.

PROGRAMS OF CONTINUING EDUCATION (PCE)

Programs of Continuing Education (PCE) offers workshops at ETS on subjects related to testing, evaluation, and measurement issues that concern and interest educators. Topics include assessing basic skills, formative and summative staff evaluation, assessment of programs in special education, and analyzing students' writing skills. PCE also produces instructional packages for use in school districts' inservice programs. Packages include content outline or texts, visual aids, background materials, and copy for preparation of handouts.

PROGRAMS FOR THE ASSESSMENT OF WRITING

Programs for the Assessment of Writing reflects ETS's long-standing interest in the teaching and evaluating of writing and offers a wide range of program services and other activities designed to help those in the academic and business communities concerned with improving and evaluating writing. Offerings include program and scoring services, such as the development of placement and descriptive tests of writing ability, and professional development workshops and seminars.

INTERNATIONAL OFFICE

The ETS International Office coordinates overseas activities of the organization and conducts programs as well as instructional, research, and consultative projects abroad. The curriculum of the instructional progam, offered yearly at ETS, includes techniques in test development and production; statistical analysis at the introductory, intermediate, and advanced levels; and program evaluation. The office also provides educational measurement and evaluation workshops and courses tailored to the needs of groups from particular countries or regions. International conferences and workshops arranged in the United States and abroad help to initiate and further collaborative research.

RESEARCH

Divisions	
Education Policy Research Institute Educational assessment; finance and governance	***Measurement, Statistics and Data Analysis Research*** Educational technology; measurement methodology; psychometric research; scientific programming; statistical research

Educational Research and Evaluation

Bicultural research; educational program evaluation; educational sociology; guidance; higher education; instructional research; occupational and career research; student development; teacher behavior

Center for Assessment and Research in Human Development

Cognition and learning; infancy; personal/cognitive development; personality and social behavior; psycho-educational development; young children

Goals

- Basic psychological investigation
- Service to the educational community
- Development and improvement of operational programs

Expenditures/1979-80

$8,225,000, including $5,690,000 from outside contractors and grantors

Projects/1979-80

Major funded projects—98 (including program-related research)

Funding Sources/1979-80

52.0%	Federal government
30.8%	ETS support
3.9%	Foundations
.2%	State and local government
13.1%	Other sources
100.0%	

Staff/1979-80

204 data analysis specialists, economists, measurement specialists, psychologists, research assistants, secretaries, sociologists, statisticians

EDUCATIONAL TESTING SERVICE

Statements of Revenue and Expenditures for the fiscal year ended June 30, 1980

Revenue		
Testing activities	$95,959,540	(90%)
Research, development, and instructional services	8,970,101	(8%)
Other revenue	1,864,689	(2%)
Total revenue	$106,794,330	(100%)

Expenditures		
Salaries, wages, and employee benefits	$57,995,762	(54%)
Professional services	11,020,305	(10%)
Printing and materials	9,921,531	(9%)
Depreciation	2,204,037	(2%)
Electronic equipment rental and services	4,805,231	(4%)
Utilities and property taxes	1,880,930	(2%)
Freight and postage/telephone and telegraph	9,262,901	(9%)
Other operating and administrative expenses	10,305,336	(10%)
Total expenditures	$107,396,033	(100%)

HENRY CHAUNCEY CONFERENCE CENTER

Who Can Use the Center

The Center, located at ETS headquarters in Princeton, is available to educational and research organizations and agencies and to other groups that sponsor conferences or meetings for educational purposes. The Center can be used for meetings that last several days or a few hours and for groups of 10 to 200.

Conference Facilities

Meeting Rooms

The auditorium barn can be used in its entirety to seat up to 200 or subdivided into separate meeting rooms. Other meeting rooms are designed for groups of 10 to 100. Facilities for film projection, slide/tape presentations, and other visual aids are available. Supporting services such as administrative and clerical assistance can be arranged in advance.

Accommodations

The Center has 100 bedrooms designed for single or double occupancy. The main dining room seats up to 200 guests.

For Information

For reservations, schedule of rates, or further information, call the operations manager, 609-921-3600, or write to the Henry Chauncey Conference Center, Box 2605, Princeton, NJ 08541.

ETS OFFICES

MAIN OFFICE

Rosedale Road
Princeton, NJ 08541
609-921-9000

William W. Turnbull, President

NATIONAL RELATIONS OFFICE

1800 Massachusetts Avenue, NW
Washington, DC 20036
202-659-0616

Alice J. Irby, Vice President

PUERTO RICO OFFICE

Suite 1115
Banco Popular Center
209 Muñoz Rivera Avenue
Hato Rey, PR 00918
809-753-6363

Protase Woodford, Director

SOUTHERN REGION

Suite 2020 250 Piedmont Avenue
Atlanta, GA 30308
404-524-4501

Scarvia B. Anderson, Senior Vice President
Charlotte Rentz, Director

Alabama
Florida
Georgia
Kentucky
Mississippi
North Carolina
South Carolina
Tennessee
Virginia

WESTERN REGION

Berkeley Office

1947 Center Street
Berkeley, CA 94704
415-849-0950

John W. Evans, Director

Alaska
Northern California
Colorado
Hawaii
Idaho
Montana
Northern Nevada
Oregon
Utah
Washington
Wyoming

Los Angeles Office

Room 216
2200 Merton Avenue
Los Angeles, CA 90041
213-254-5236

J. Richard Harsh, Director

Arizona
Southern California
Southern Nevada

Southwestern Office

Suite 100
Southwest Tower
211 East Seventh Street
Austin, TX 78701
512-478-8191

Reginald A. Corder, Director

Arkansas
Louisiana
New Mexico
Oklahoma
Texas

MIDWESTERN REGION

1 American Plaza
Evanston, IL 60201
312-869-7700

Michael Kean, Director

Illinois
Indiana
Iowa
Kansas
Michigan
Minnesota
Missouri
Nebraska
North Dakota
Ohio
South Dakota
West Virginia
Wisconsin

NORTHEASTERN REGION

111 Washington Street
Brookline, MA 02146
617-739-2210

George W. Elford, Director

Connecticut
Delaware
Maine
Maryland
Massachusetts
New Hampshire
New Jersey
New York
Pennsylvania
Rhode Island
Vermont

For further information regarding ETS programs and activities, contact the Information Division, Educational Testing Service, Princeton, NJ 08541/609-921-9000. The cable address is EDUCTESTSVC; telex 84-3420. *Educational Testing Service is an Equal Opportunity Employer.*

SIDE-BY-SIDE COMPARISON OF H.R. 1662 PROVISIONS AND CURRENT PRACTICES OF GUIDANCE AND ADMISSIONS TESTS ADMINISTERED BY ETS

Major Points:

- Admissions test sponsors served by ETS already provide **voluntarily** most of the information about tests required by H.R. 1662. For example they:

  —provide information about each test, its purposes, its scoring, and the interpretation and use of scores, to test-takers and institutions,

  —publish studies about their tests,

  —voluntarily disclose **nationwide** the questions on several major administrations of each admissions testing program each year.

- H.R. 1662 would require the establishment of a bureaucratic entity in the Department of Education to collect information, most of which is already publicly available, and to prepare a report to Congress on subjects which have already been widely and thoroughly debated.

- This bill's forced disclosure of confidential documents governing business transactions between non-profit organizations is unprecedented. It represents unwarranted governmental intrusion into private business affairs.

KEY PROVISIONS OF H.R. 1662
(The Educational Testing Act of 1981)

Sec. 3 Information to Test Subjects and Postsecondary Education Institutions

- Test agencies must include in registration materials information as to:

—test's purpose[1]

—subject matter and knowledge and skills measured[2]

—interpretation of test results, including test explanation, correlation between score and future school success, standard error of measurement and correlation between test score and career success[3]

—effects on and uses of test scores, including extent of improved accuracy in prediction over Grade Point Average, comparison of scores according to income, and extent of score improvement from coaching[4]

—form of score report, alteration of scores, and use in prediction[5]

—accuracy of scoring, timeliness of reporting, and privacy[6]

—students' property interest in test results, retention, storage and future use of scores[7]

—time period within which students and institutions will receive scores[8]

—special services for handicapped students[9]

—test disclosure, test privacy, and appeal procedures[10]

- institutions which are test score recipients must also receive above information[11]

- Students and institutions must be notified if scores are delayed ten days[12]

CURRENT PRACTICE OF GUIDANCE AND ADMISSIONS TESTS ADMINISTERED BY ETS (SAT, PSAT, GRE, LSAT, GMAT, TOEFL)

Sec. 3 Information to Test Subjects and Postsecondary Education Institutions

1 Provided by all tests in registration materials.

2 Provided by all tests in registration materials.

3 All test programs provide, either in registration materials or materials distributed at the time of score reporting, information on how to interpret test results, including the standard error of measurement, relationship of scores to first year grades, and the use of scores in conjunction with grades in school. Most programs include statements to the effect that the tests are not designed to predict future success in a given field and, consequently, do not include statements about how well scores predict success or income.

4 See 3 above. This information is available in published research reports for some programs; others do not collect income information from candidates. ETS and test sponsors believe it is not desirable to report this information in preregistration materials for candidates because it could be misleading since individual candidates in all income groups can and do obtain scores at all levels.

Meaningful statistical averages or percentages on the effects of coaching are not available because the results of those studies that have been done vary so widely. Reports of all such studies are publicly available. All programs include general statements on the effects of coaching in their registration or test preparation materials (except TOEFL which is an achievement test).

5 Provided, where applicable, in registration materials and/or with score reports.

6 Provided by all tests in registration materials.

7 "Property interest" is an inappropriate reference; however, the issue of confidentiality with respect to retention, storage and future use of scores is discussed in all registration materials.

8 Provided by all tests in program materials.

9 Provided by all tests in program materials.

10 Provided by all tests in program materials.

11 Institutions are provided the same information provided to test-takers.

12 Those testing programs which have experienced delays in score reporting have routinely notified students and institutions at the time of regular reporting that a score or scores will be delayed.

KEY PROVISIONS OF H.R. 1662

Sec. 4 Reports and Statistical Data and Other Information

- The Secretary must be provided the following information:

  —any study, evaluation or statistical report about a test which a test agency prepares or causes to be prepared, or provides data for[1]

  —a copy of service contracts between test developer and test sponsoring agencies[2]

- All documents so submitted shall be considered records subject to the Freedom of Information Act[3]

- The Secretary must report to Congress on relationships between test scores and income, race, sex, ethnicity, handicapped status, and completion of test preparation courses[4]

Sec. 5 Promoting a Better Understanding of Tests

- Test agencies must file with the Secretary and send test subjects, upon request:[1]

  —a copy of test questions used in determining raw scores

  —the correct answers (and to test subjects, their individual answer sheets)

  —a statement of rules for transferring raw scores (and to test subjects, the raw score used to calculate the final score)

- Test agencies may charge a nominal fee for this service[2]

- These provisions do not apply to tests administered to less than 5,000 per year[3]

- Documents submitted to the Secretary are considered records subject to the Freedom of Information Act.[4]

CURRENT PRACTICE OF GUIDANCE AND ADMISSIONS TESTS ADMINISTERED BY ETS

Sec. 4 Reports and Statistical Data and Other Information

1 Routinely ETS publishes studies it carries out, if not identifiable with particular individuals or institutions, or when not precluded from release by contract agreement. However, ETS and test sponsors cannot be responsible for studies conducted by others based on data provided by them.

2 ETS and its clients follow the typical business practice of keeping their contractual agreements confidential.

3 See 1 and 2 above.

4 Routinely ETS publishes research findings on these subjects where not precluded because of confidentiality considerations or contractual agreement.

Sec. 5 Promoting a Better Understanding of Tests

1 All admissions test sponsors served by ETS have instituted **voluntary** procedures for disclosing their tests **nationwide**:

SAT: Four of six regular national administrations plus a special October administration in six states are disclosed.

LSAT: All four major administrations are disclosed.

GRE: Two of the five major administrations of the aptitude test are disclosed.*

GMAT: All four major administrations are disclosed.

TOEFL: Five of twelve administrations are disclosed.

PSAT: Both of the two administrations are disclo

For all testing programs there is at least one opportunity for a Saturday Sabbath observer to take a disclosed test.

*(Note: Constraints resulting from introducing a newly formatted test in Oct. '81 prevents disclosure of additional GRE test forms in the 1981-82 testing year. Three are planned for 1982-83.)

2 Nominal fees are charged for disclosure.

3 See 1 above.

4 Protecting copyright on disclosed tests is a concern of ETS and its test sponsors. A perception by the public that material that is "available" is not protected by copyright could lead to inappropriate commercial uses of test material and might also jeopardize the security of other undisclosed test material.

KEY PROVISIONS OF H.R. 1662

Sec. 6 Privacy of Test Scores

- Test scores may not be released to any party unless authorized by test subject[1]
- Test agencies may release all previous scores to designated score recipients[2]

Sec. 7 Testing Costs and Fees to Students

- Testing agencies must report to the Secretary

—the closing date of their testing year

—the number of times a test was taken during the testing year

—the number of test-takers taking the test once, twice, and more each year

—the number of refunds given to registrants who didn't take the test

—the number of test-takers for whom the test fee was waived or reduced

—total amount of fees received for each test program per year

—total revenue received from each program

—expenses to the test agency including test development and overhead costs

—number registering for admissions data assembly services and for score reporting services for which fees are charged

—amount of revenue received for admissions data assembly and score reporting services

—expenses for admissions data assembly and score reporting services.

CURRENT PRACTICE OF GUIDANCE AND ADMISSIONS TESTS ADMINISTERED BY ETS

Sec. 6 Privacy of Test Scores

1 This is existing practice.

2 This is existing practice.

Sec. 7 Testing Costs and Fees to Students

Much of the information regarding test repeaters, refunds, and fee waivers, is maintained by ETS but not regularly reported because of lack of public interest. Information about ETS fees, revenues, and expenses, except that which is considered confidential according to standard business practices, is publicly available. Information concerning data assembly services is not an ETS responsibility, but rather one of the Law School Admissions Service.

Mr. Weiss. Another area that we had some dispute over is a question as to what the test developer or test giver ought to be doing to bring to the attention of the test taker, that the information included in H.R. 1662 is available to him or her. There have been different approaches to that. The law school people have a box on the top sheet of their test which the test taker simply has to check if in fact he or she wants the results. It's not even going to be necessary as of, I understand, this year or next year because everybody's going to get them, but that was their approach to them. The ETS approach was to have a separate sheet of paper which the test taker had to locate within the packet and then having located it, to make some notation on it, and then, I guess, submit it in some way to ETS. The question, I think, I asked of a witness from the College Board at the last hearing and I ask it of you, why can't you make it as simple as the law school admission people made it.

Mr. Anrig. Mr. Chairman, because you made an error that I used to make all the time until I came to ETS and that is you assumed that the College Entrance Examination Board and the Educational Testing Service are one in the same—they are not. Educational Testing Service is contracted by the College Board, by the Law School Admissions Board, by the Graduate Management Advisory Council, and by the Graduate Records Examination Board, to perform certain services. The decisions on those programs are made by each of those independent governing boards. We administer whatever decisions those are that are carried out. Obviously we try to influence those decisions and advise on them, but the decision rests with the separate governing boards which, in turn, contract with us to carry out their decision.

My own feeling is if you're going to release it, put it out there. I have no problem at all with that, but that's not my decision. The decision rests with the governing board. In fact, I think that's not as big an issue as has been made of it. We're talking here about youngsters who are going on to college. The difference of them dealing with a checkoff point or filling out a separate form, they are experts in filling out separate forms. They've been doing it all their lives, and many of those forms I required as a state commissioner. So I don't think that's a big issue.

Mr. Weiss. I think it is, because you know what happens. We get the witnesses coming back to us telling us how unnecessary this legislation is, arguing that the test takers don't want the results, right? Now, it seems to me you can't really have it both ways. You can't hide the request form—I'm not talking about you, whoever is giving the test at that time—can't hide the request form and then argue that because it's not found or it's not sent in in large enough numbers that students are not interested in submitting a request form.

Mr. Anrig. Well, two points, Mr. Chairman. First, you didn't hear me today advance the argument on how many people are asking or not asking for the forms. I don't think that's a worthwhile argument.

Mr. Weiss. Yes, but the College Examination Board people made it.

Mr. Anrig. That's their perogative to do so. I can only speak for Educational Testing Service, and you did not hear that referred to in my testimony nor do you see it other than in gross numbers in the appendices to the testimony. I don't think that's an argument we should get into, but that's my judgment and others are entitled to theirs. The second point is the forms have not been heavy. The forms are there, I've seen them. They're not either complicated or hidden. Is there a better way to do it? In my judgment, yes. I happen to like what the Law School Admissions Council is doing, but it's not my judgment. The judgment rests with the boards that govern those test programs.

I think what is happening and I believe the July hearing inferred this was that over time, people are going to be more alike than different. I think you said that a couple of years from now you're going to brag that you just had a great innovation by releasing something like that and really you could have done it all along. I was sympathetic to that testimony, but nontheless I don't call those shots. That's for somebody else to call. I don't think the present system is cumbersome for college-bound students. I saw nothing in there that was either hidden or unduly bureaucratic or difficult for a student to complete. It is different and I think it would be nice if we had less of that difference over time and I think we will.

Mr. Weiss. I think that given the interrelationship between ETS and the College Entrance Examination Board, it is probably fair to conclude that if ETS were to suggest how they do it that they might at least get a sympathetic hearing. They may not follow it but I think they would listen to you sympathetically, wouldn't you think so?

Mr. Anrig. Let me put it this way, I would hope so and I'll certainly make that as a result of our discussion here today, Mr. Chairman, I'll bring that comment back to them. But let me assure you the sense of independence between the College Board and ETS is almost as much as the sense of independence of New York State versus New Jersey State.

Mr. Weiss. Well, I won't comment on that. Another area of some dispute and I think you addressed it to some extent, is that we've had testimony in the course of the last couple of years from scholars in the field who told us how much difficulty they had in getting both in-house reports that ETS had worked up or had commissioned and/or some of the raw data on which those reports were based.

Now, you were suggesting I think in the course of your testimony that there is going to be a change in that attitude. Some of the witnesses we had told us that they had to go to court to try to get some of that material. Question, how is it going to work in the future?

Mr. Anrig. Mr. Chairman, that is an issue which the ETS board of trustees had assigned to its Subcommittee on Research and Development because it wants to resolve that issue. In fact, the past history has been that as far as I've been able to track it back and I admit to being new at this that requests for information generally were acceded to if the information were available. Some of the requests that came in were not for raw data, as I understand it, but for data that was worked, that is to request that something be done about the data and then fulfill the request. Well, that has costs connected to it. We're not in the business of being a research library for researchers in terms of asking us to do something pro bono. The raw data, that is the basic data, and the data to the extent that we have worked it for reasons of our own planning and evaluation will be put on what are called on public-use tapes. Those public-use tapes will be available either through ETS or if we can find a third-party kind of way of doing that we're looking at that issue now, and that will be available at cost or very minimal kind of cost for whatever that service is. My general feeling as a rule of thumb is that if we have it we ought to let others have it, and that the burden of proof on why not to do so should fall on us internally and I've asked my own staff to approach it that way. If we've got it, let's put it out there. If there's some sound reason proprietary in nature or copyright in nature or certainly on the issue of individual privacy where we should not do that, then let's prove to ourselves why that should be the prevailing decision on that. So our basic approach will be that way. Those recommendations were voted upon yesterday in the committee. They'll be brought to the board of trustees at the April meeting of our board and I anticipate very speedy action on that and I will be in support of that action.

Mr. WEISS. The other area you touched on along those lines deals with test taker accessibility, complaints, grievances, and so on. Can you spell that out? You said that you've got somebody who's been specially assigned to deal with that because the testimony that we've had up to this point indicates that yes, ultimately if somebody is fortunate enough and persevering enough that maybe they will get a remedy. But there's no process at all as to how to start. You send a complaint off to somebody or to ETS and it's just as likely to be totally ignored as it is to be responded to.

How are complaints either as to scoring, grievances as to the ambience or the setting in which the examination was given, or whatever the various complaints, how are they going to be handled?

Mr. ANRIG. Mr. Chairman, first of all the position that's been advanced as you've just described it is not either true or fair. I believe you've been down at ETS and visited our Ewing site where people can call in directly and they get the answer right on the phone. What's my score? Was it sent to such-and-such a school? What did I say on this particular part of the questionnaire? Those answers are given by phone, calling into ETS, no matter where you call in it gets switched over to that Ewing site, and there's a room full of people doing that. I think you saw that during your visit. If you didn't, come on back and we'll take you over there and have you see it.

ETS works very hard at responding to those student complaints. The problem I found as I looked into that very same question because I feel strongly about the accessability about an organization including to those who want to complain about it was that a student who once took the SAT complained one time one way and then as he or she moves to graduate school for the graduate record exam there's a different process, and if they want to go to law school there's a different process. That's confusing to our consumers. So what I want to have is one way, not on a centralized basis, but one procedure that if you want to complain about anything at

ETS whether in a regional office or calling into the main switchboard or writing me, that it will be handled in the uniform way in terms of being received, logged, assigned to a person, resolved, and back to the complainant within a period of time that we set in advance. So that's something we're going to have in place. I put the student concerns coordinator in charge of getting that in place. We will have it fully operational by January 1, and I think it's a problem I've identified and we can resolve rather quickly.

Mr. **Weiss**. We had a witness who took the examination and the monitor concluded, surmised, perhaps suspicions vibrations, that the test taker had done something improper or unethical. The monitor made that notation, and the student was then given a failing grade, or low grade, or no grade. I don't recall specifically on that. A teacher-instructor who knew that student very well spoke to the student, in fact then spoke to the monitor, and obviously there was nothing objective about the conclusion or determination of that monitor. They had not, although at that point it was over 1-year-old, they had not been able to arouse the attention of anybody at ETS to the injustice that seemingly was done just as a matter of process. Never mind the merits of the case, I don't know what the merits are as far as the process was concerned, and I wonder what you're doing or what your thoughts are about having a formalized grievance procedure so that you don't have some personnel in either ETS or the college entrance examination board or whoever making what in essense are life and death decisions without the right of due process.

Mr. **Anrig**. Mr. Chairman, a former Governor of New York State used to start off his speeches by saying, "Let's get the record straight," Al Smith. Let's get the record straight on this issue. I used to be a principal and every time a student was sent to my office by a teacher for causing trouble in the classroom when I asked that student what happened I got one story. When I found out what happened from others involved there was a different perspective on it. One of the problems of a hearing process such as you're in in this situation is that a person comes before you and makes an allegation. If there's not the opportunity to rebut that allegation on the facts of that, then it stands on the record.

Let me talk about our procedures because I can't answer that particular case, I wasn't here at the time. But our procedures are such that if a question of test security is what the policy that apply here is raised, that is sent to a unit which deals with that very issue and specializes in it. No single piece of evidence alone is enough to establish that there's been a breach of security. There must be a combination of evidence which is compelling, and that evidence and those criteria are spelled out and published and I'd be glad to submit them for the record. There is an appeal procedure from that. There is nobody that is given a failing grade. They simply are not given a score and they're notified that they will not receive the score. They may appeal, or they may take the test over. If they take the test over I believe if I remember correctly that's done without charge. So the opportunity is given to take the test again without charge and also there is an appeal route. The reason I know this is I've already received an appeal from that procedure and that appeal eventually comes to the president of the organization. There's a very careful process. When I first heard that by computer scanning we identified these problems I was scared to death. As I got into the extraordinary combination of circumstances that must be present for a test security allegation to be made, I didn't have that concern because there must be established clear patterns beyond the individual instance of the test supervisor in the center. In other words they would have looked also at the pattern of test answers. And if you find similarity to the person sitting next to them, that kind of thing, then you do begin to get into evidence.

So the problem you had in that instance, I believe, was the student coming into the principal's office and telling his or her side of what went on back in the classroom. In fact, the policy of ETS in that area are in print, are very clear, do include an appeals procedure, and I'd be glad to submit that for the record.

Mr. Weiss. How long has that been in place?

Mr. Solomon. Several years.

Mr. Anrig. Mr. Robert Solomon, the executive vice president of Educational Testing Service says it's for several years.

Mr. Weiss. There's been no change in that process for the last few years?

Mr. Anrig. There has been change, but I'm not familiar with it.

Mr. Weiss. Would you outline the change for me?

Mr. Anrig. Pardon me?

Mr. Weiss. Would Mr. Solomon?

Mr. Anrig. With the Chair's position I'll call on Mr. Solomon.

Mr. Weiss. Outline what changes have taken place in the last couple of years.

Mr. Solomon. Mr. Chairman, the more recent change is that in two programs, first with the law school admission test and more recently with the college board scholastic aptitude test, the appeal procedure now provides for going if the student requests to arbitration provided by the American Arbitration Association.

Mr. Weiss. When was that?

Mr. Solomon. The law school instituted that procedure at least 5 or 6 years ago, I don't recall exactly; and the college board instituted that procedure this past year. The other programs for which we administer tests are also considering the arbitration appeal route as well. Previous to that was an appeal procedure but the outside arbitrator was not part of that appeal procedure.

Mr. Weiss. Thank you very much. So no longer will ETS or CEEB be in fact the judge, jury, prosecutor, the whole works in the situation.

Mr. Solomon. We try never to be that, sir, but we have included the arbitration procedure now.

Mr. Weiss. All right, thank you very much. Then, I guess, the last question I have is a factual one. How many ETS sponsored tests are given annually?

Mr. Anrig. In terms of numbers or the titles of tests?

Mr. Weiss. No, numbers. How many SAT's for example?

Mr. Anrig. With the Chair's permission, I'll submit that for the record. I don't know that offhand. It's somewhere around 1.2 million. I can tell you what it is, I've got a document here for 1981 which is not quite the current year, but the admissions testing program, the SAT goes to 1,255,000 students. The total admissions testing program, which would include the achievement test, the SAT, is 1,527,000. That was for the years 1979–80.

Mr. **Weiss**. Thank you very much. Would you make sure that whatever additional materials you will be submitting are submitted within 10 days so the record can be closed within a reasonable period of time?

[The information referred to follows:]

RR-81-19

RESEARCH

REPORT

ISSUES OF EFFECTIVENESS AND EQUITY IN THE COACHING CONTROVERSY: IMPLICATIONS FOR EDUCATIONAL AND TESTING PRACTICE

Samuel Messick

Educational Testing Service
Princeton, New Jersey
August 1981

Issues of Effectiveness and Equity in the Coaching Controversy: Implications for Educational and Testing Practice

Samuel Messick

Educational Testing Service
Princeton, New Jersey
August 1981

Issues of Effectiveness and Equity in the Coaching Controversy: Implications for Educational and Testing Practice

Samuel Messick

Educational Testing Service

ABSTRACT

Multiple sources of controversy over coaching for scholastic aptitude tests are addressed. These include disputes over the meaning of scholastic aptitude, over the meaning of coaching, over the nature of the fundamental research questions, over the adequacy of the empirical evidence for coaching effectiveness, over the possible implications of effective coaching for student performance and for test validity, and over the consequent ethical imperatives for educational and testing practice. Three possible outcomes of coaching having different implications for college performance are discussed -- genuine improvements in the abilities measured by the test resulting in commensurate increases in test scores; enhanced test-taking sophistication (item familiarization, guessing, pacing) resulting in increased test scores that are more accurate assessments of ability; and, heightened test-taking artifice (stratagems and answer-selection tricks) resulting in increased test scores that are inaccurately high assessments of ability. The first two if realized would be beneficial from the standpoint of both test validity and student performance, while the third is only incidental with professionally made tests. A framework for gauging the practical utility of coaching effects is presented and several ostensible yardsticks of coaching impact are discounted as largely irrelevant.

Issues of Effectiveness and Equity in the Coaching Controversy: Implications for Educational and Testing Practice[1]

Samuel Messick

Educational Testing Service

The controversy over whether or not coaching works for tests of scholastic aptitude such as the College Board SAT is in actuality a multitude of controversies -- over the meaning of scholastic aptitude, over the meaning of coaching, over the nature of the fundamental research questions, over the adequacy of the empirical evidence for coaching effectiveness, over the possible implications of effective coaching for student performance and for test validity, and over the consequent ethical imperatives for educational and testing practice. Since misconceptions and misunderstandings abound on each and every one of these points, discussions of the coachability of scholastic aptitude tests readily reduce to rhetorical and political posturing when what is needed is more extensive rational analysis and scientific appraisal of research evidence. And as we shall see, since much of the controversy reflects poorly articulated ethical stances, what is also needed is a clearer conception of the values underlying rival ethical imperatives.

In any event, it is not so much a question of eschewing rhetoric, because in essence test "validation is a rhetorical process" (Cronbach, 1980, p. 102). Rather, it is a question of ensuring that the rhetorical arguments reflect facts and rationales as well as values and beliefs. Granted that "empirical evidence, task analysis, formal assumptions, everyday beliefs, and valuations are intertwined in the argument that supports a test use" (Cronbach, 1980, p. 101; Messick, 1980b, 1981b), nonetheless the role of evidence should be central and directive in tempering the trade-offs embodied in our current best scientific judgment about a value-laden issue such as coaching. To be sure, since it rarely completely dispels divergent interpretations, evidence -- that is, "both data, or facts, and the rationale or arguments that cement those facts into a justification of...inferences" (Messick, 1980b, p. 1014) -- may not fully determine the outcome or conclusion. Evidence does serve, however, to limit the range of plausible alternatives. In an attempt to reinforce the directive role of evidence in current arguments over coaching, the present article summarizes research findings from studies of coaching for the SAT in the context of a conceptual analysis of the meaning and likely import of the coachability of such ability tests. The intention is to illuminate if not resolve, in the light of recent research results, the central issues of meaning, effectiveness, and equity of access inherent in the coaching controversy.

Controversies Over Meaning

As so often happens in continuing controversies, the key terms -- in this case "coaching" and "scholastic aptitude" -- take on different meanings for different proponents. Multiple meanings of test constructs often lead to surplus or excess meanings that go beyond defensible score interpretations, and "excess meanings lead to misuses and misinterpretations of tests" (Anastasi, 1980, p. 1). Imputed meanings for scholastic aptitude range all the way from "fixed endowment," which implies that valid scholastic aptitude tests should be essentially uncoachable, to "direct learning outcome," which implies that they should be as responsive to coaching or instruction as any other measure of educational achievement such as tests of history or biology. But scholastic aptitude as measured by the SAT falls at neither of these two extremes: The SAT measures developed -- or more precisely, developing -- abilities which are by no means fixed but which by the senior year of high school are relatively stable in their development. This implies that the SAT may indeed be responsive to coaching or instruction to some degree, but it should not be nearly as responsive as the typical educational achievement test of subject-matter attainment.

Similarly, imputed meanings for coaching range all the way from short-term cramming and simple practice on sample items at one extreme to long-term instruction aimed at knowledge and skill development at the other. It is here maintained that any intervention program designed to improve SAT scores may be considered an instance of coaching, wherever it may fall along this continuum of meaning, but that such "coaching" programs should not be averaged together in a single conglomerate category because different program characteristics, such as duration or curriculum emphasis on skill development, may be associated with differential SAT score effects (Messick, 1980a).

Scholastic Aptitude as Developed Ability

One important source of controversy over the meaning of scholastic aptitude derives from differences in the popular and technical uses of the term "aptitude." The dictionary (Webster, 1969) denotes aptitude as "capacity for learning...[or] natural ability" (p. 44), "a natural liking for some activity and the likelihood of success in it" (p. 352). In the loose language of popular usage, it seems only a small slip to move from "natural ability" to "innate ability" and from there to "fixed ability." But it is a slip nonetheless and, from the standpoint of its putative implications for aptitude measurement, an enormous one. For if aptitude constitutes fixed ability, then valid aptitude tests should presumably not be coachable. Any such tests that in fact prove to be coachable are not valid measures of aptitude by this chain of reasoning, but rather assess general educational achievement; they should therefore be treated in selection decisions on the same basis as other measures of achievement. If purported aptitude tests are interpreted as being more fundamental or more instrumental in predicting future success than are typical achievement tests and if those so-called aptitude tests are indeed coachable, then the issue comes to be cast in terms of fair test use: By this line of argument, coachable aptitude tests are ipso facto unfair to students not having access to high quality coaching programs. But the equity issue is much more complicated than this even granting the foregoing simplistic premises, because coaching is a form of instruction and coachable aptitude tests are said to be akin to achievement tests, which are expected to be responsive to instruction -- in this involuted sense of unfairness where tests bear the burden for societal inequities, educational achievement tests have always been unfair to students not having access to high quality _instructional_ programs.

In contrast to nontechnical dictionaries, a compendium of psychological terms (English & English, 1958) asserts that "APTITUDE (which formerly carried implications of _innateness_) has now been specialized in technical writing to refer to the fact that the individual can be brought by a specified amount of training to a specified level of ability, either general or special" (p. 1). Furthermore, "an APTITUDE TEST is merely one form of ability test,...a measure of present characteristics that has been found to be predictive of capacity to learn" (p. 39) or, more generally, predictive of future performance. Along these same lines, two closely related but distinct features of aptitude are noted by Cronbach and Snow (1977): Generally speaking, aptitude is "any characteristic of a person that forecasts his probability of success under a given treatment" (p. 6) -- that is, aptitude is a forecaster of learning or performance; speaking "psychologically, aptitude is whatever makes a person ready to learn rapidly in a particular situation" (p. 107) -- that is, aptitude is a facilitator of learning or performance. Both of these connotations are inherent in the typical rationales for interpretation and use of scholastic aptitude tests.

Scholastic aptitude tests such as the SAT are not direct measures of innate intelligence or fixed endowment, nor are they measures of subject-matter attainment as in the usual educational achievement test. The SAT was explicitly designed to differ from achievement tests in subject-matter fields in the sense that its content, being derived from a wide variety of substantive areas, is not tied to a particular field of study, curriculum, or program. The SAT measures developed abilities of verbal and mathematical reasoning and comprehension that are acquired gradually over many years of experience and use in both school and nonschool settings, being exercised to some degree in all subject-matter areas at all levels of schooling as well as in response to real-life situations. According to Carroll (1978), "to the extent that tests of the SAT type are valid in predicting college or graduate-school success, it is undoubtedly because they provide a good indication of the extent to which applicants have at the time of testing developed or acquired, and can exhibit through their performance on a test, certain general intellectual skills in handling verbal, quantitative, and symbolic information that are contributory or even necessary to high-level success in academic studies" (p. 78). In other words, scholastic aptitude tests are general forecasters of academic learning and performance because they measure general facilitators of academic learning and performance.

A critical feature of this technical formulation of the construct of aptitude is that scholastic aptitudes are viewed as developed abilities, not fixed abilities, so one might expect high quality instruction over an extended period of time to improve them. But since these general intellective skills develop gradually over a number of years as a result of everyday experience as well as formal education, they may be relatively difficult to enhance markedly in late adolescence through brief courses of intervention. This is in contradistinction to the specific knowledge and skills tapped by typical educational achievement tests, which should be relatively responsive to high quality instruction even in the short run. As Carroll (1978) put it, "a low score on...a [scholastic aptitude] test is no guarantee that the individual cannot acquire, during a subsequent period, the skills and abilities that are tested, but it is an indication, at least, of the probability that the individual could acquire these skills and abilities, if at all, only with much expense of time and effort and with careful instruction" (p. 78).

Repeatedly throughout this discussion, contrasts have been drawn between scholastic ability tests on the one hand and educational achievement tests on the other. In idealized form these two types of tests define opposite poles of a continuum, with most actual educational tests -- by virtue of sharing various features of each pole -- falling somewhere in the intermediate range. For example, scholastic ability tests measure general intellective skills such as verbal reasoning by means of diverse content drawn from a multiplicity of topical areas, whereas educational achievement tests measure specialized knowlege and skill using relatively circumscribed content drawn from a particular subject-matter field. Yet an achievement test in political science, say, might include items to assess skill in reasoning with political concepts; to the extent that such specific reasoning also reflects more general reasoning abilities, the test might prove to be more broadly predictive beyond political science to related fields.

In many ways, an educational achievement test is akin to a job-sample test -- that is, a test in which the items present task and stimulus conditions in roughly the same form and in the same relationships as they appear on the job (or in the curriculum), thereby eliciting responses representative of behaviors similar to those on the job (or in the course of academic learning). Scholastic ability tests, in contrast, measure reasoning and comprehension skills that contribute to successful performance in a number of important academic jobs. Although such ability tests might not be as predictive as the job-sample test for the specific job (or field of academic learning), they are more generally predictive across a variety of jobs or fields calling for those reasoning skills (Messick, 1981a).

Among other things, the two types of tests differ in what Snow (1974; Coan, 1964) calls referent generality -- that is, in the breadth or variety of performances presumed to be covered or interpreted by the respectively measured constructs of educational achievement and scholastic ability. One implication of this difference in referent generality, according to Snow (1980), is that "if one chooses to work with a relatively specific level of achievement measurement, the best predictor...will be relevant prior achievement, [and] next best will be general scholastic ability...If one chooses to work at a more general level of achievement construct, [such as college grade-point average,] the best...measures for prediction purposes will be general scholastic ability, followed by specific prior achievement tests" (Snow, 1980, p. 46).

Coaching as Instruction, Demonstration, and Practice

Just as there is a progression or ordering of educational tests ranging from measures of scholastic abilities at one pole to measures of subject-matter attainment at the other, so there is a progression of types of preparation for taking examinations ranging from practice on sample tests at one extreme to intensive instruction aimed at developing knowledge and skill at the other. Historically, coaching for scholastic ability tests has tended to fall toward the practice side of this continuum, emphasizing test familiarization and test-taking strategies, while coaching for achievement tests has tended to fall toward the instructional side, emphasizing directed study and review of subject-matter content. This wide array of possible forms of coaching has led to heated controversy over its meaning, because different parties often selectively restrict their referents to different portions of the continuum. At the same time, the full range of coaching possibilities has come to be applied to any type of educational test, thereby making preferential usage appear not only restrictive but arbitrary. For example, in regard to coaching for the SAT, some writers have limited their use of the term to short-term drill, test practice, and test-wiseness training (cf. Pike, 1978), whereas others include under the same rubric virtually full-time instruction at specialized preparatory schools for periods of six months or more (cf. Slack & Porter, 1980). Adherents of the former position underscore a distinction between coaching and instruction, whereas adherents of the latter position view such differences as immaterial, implying that anything resulting in improved test scores testifies to coachability.

In an effort to avoid disputes over what is or is not coaching, we will here accept as coaching any intervention procedure specifically undertaken to improve test scores, whether by improving the skills measured by the test or by improving the skills for taking the test, or both. Thus, for our purposes coaching may fall anywhere in the broad range bounded by the two extremes of practice and instruction, embracing any combination of test familiarization, drill-and-practice with feedback, motivational enhancement, training in strategies for specific item formats and for general test taking (including advice on pacing, guessing, and managing test anxiety), subject-matter tuition and review, and skill-development exercises. In this view, coaching for educational tests is any procedure specifically oriented toward the improvement of test scores as distinct from nontest-specific learning experiences and cognitive growth -- which may also result in improved test scores. This broadly inclusive stance is consistent with the dictionary definition of the word coaching, which is "to train intensively by instruction, demonstration, and practice" (Webster, 1969, p. 158).

Controversies Over Effectiveness

Even if we were to have agreement on the meaning of scholastic aptitude as developed ability and on the meaning of coaching as any intervention procedure for improving test scores, there would still be ample controversy over the effectiveness of coaching and, indeed, over the way in which the question of effectiveness is posed. As Kaplan (1964) maintains, "How we put the question reflects our values on the one hand, and on the other hand helps determine the answer we get" (p. 385). The challenge is to have sufficient recourse to relevant research evidence so that our values do not solely determine the answer we get. The problem is that research evidence is selectively sought in the context of the question posed so that facts and values are intimately intertwined in our efforts at interpretation. One safeguard is to attempt to take values explicitly into account throughout the research endeavor. Another is to pose the research question in alternative ways so that data are collected and analyzed from alternative perspectives -- in the hope that convergence of evidence would lead to consensus of interpretation and that divergence of evidence would lead to exposure and examination of value differences (Churchman, 1971, 1979; Messick, 1975, 1980b).

Yes-No Questions Versus Questions of Degree

Studies of the effectiveness of coaching for the SAT have consistently posed the research question in categorical form, seeking to ascertain in general terms whether or not coaching works and, if it does, to determine the overall amount of score increment attainable through coaching (Messick, 1980a). Those viewing the coachability of scholastic aptitude tests as primarily an issue of fair testing practice tend to phrase the question in this way. For them, coachability implies unfairness if student access to effective coaching is not equitable or, at the least, if examinees are not uniformly apprised of the kinds of special preparation they should undertake. However, if coachability is viewed not only as an issue of fair testing practice but also as an issue in the construct validity of scholastic aptitude tests as measures of developing abilities, then the research question is more aptly posed as one of degree: It is not just a question of whether coaching works or not, but of how much student time and effort devoted to what kinds of coaching experiences yield what level of score improvements. From this standpoint, it matters to what _degree_ coaching is effective and by what _means_, for although developed scholastic abilities may be further developed by relatively long-term programs emphasizing comprehension and reasoning skills, they should not be readily improved by short-term programs stressing test-taking strategies or drill-and-practice. This is not to say, however, that programs of the latter type might not lead to some score improvement as a consequence of enhanced test wiseness or reduced anxiety about what to expect.

The first viewpoint, by virtue of focussing simplistically on whether coaching works or not, leads to a simple tallying of statistically significant and nonsignificant coaching effects or to an averaging of attained coaching effects across available studies (Slack & Porter, 1980). The second viewpoint, by virtue of focussing on the relationship between student time and effort devoted to coaching and the size of associated score effects, leads to the calculation of correlation coefficients across coaching programs between estimated coaching effects and such indices of student effort as program contact time (Messick & Jungeblut, 1981). Each of these viewpoints thus provides a distinctive organizing framework for summarizing the research findings from studies of coaching for the SAT, which is our next topic for detailed examination. One might hope that such a juxtaposition of alternative ways of integrating research findings would provide some closure on the issue of coaching effectiveness -- and it does -- but the situation is beclouded by a variety of flaws in experimental design that variously distort the extant studies. This leads to yet another controversy, this time over the adequacy of the empirical evidence for coaching effectiveness.

Categorical Answers Versus Relational Answers

Whether posed in categorical or in relational terms, the question of coaching effectiveness must perforce be answered in comparative terms by appraising obtained score gains associated with coaching against the baseline of experiential growth in ability that may occur over the same time period in the absence of any coaching program. Since developed scholastic abilities continue to develop during the high school years in response to both formal schooling and nonschool experiences, any score increases exhibited by groups of coached students need to be evaluated in comparison with the score gains of equivalent control groups of uncoached students. A major source of discord in summarizing the import of research results on coaching is that several of the studies either utilized no control group at all or else employed control groups that were seriously nonequivalent to the treatment groups receiving coaching, thereby rendering highly questionable both the size and the meaning of obtained score effects. This is particularly unfortunate because those studies having no control groups at all happened also to exhibit the largest score increases of any studies to date and to entail coaching programs that were by far the longest in duration (Pallone, 1961; Marron, 1965; Coffman & Parry, 1967), while among those studies having patently noncomparable control groups is one of the largest and most ambitious coaching investigations yet attempted (Federal Trade Commission, 1979).

In the former case, there is no really satisfactory way in the absence of control groups to adjust the obtained score increases for expected experiential growth in ability, although various adjustments based on comparisons with national norm groups or with control groups from similar studies have been attempted with debatable credibility (Pike, 1978; Slack & Porter, 1980; Messick & Jungeblut, 1981). In the latter case of nonequivalent control groups, regression techniques may be used to control statistically for those group differences for which some index or measure is available, but there is no way of adjusting for any unmeasured personal characteristics that might have influenced both the student's participation in the coaching program and that program's apparent effectiveness.

This is the problem of selection bias, which is a pervasive one in studies of coaching. It arises whenever systematic differences that are correlated with the dependent variable -- in this case with SAT performance -- exist between the experimental and control samples. Systematic differences resulting from student choice of the treatment rather than from experimenter choice of the student are called self-selection bias. To be sure, those group differences for which reliable measures are available may be effectively adjusted statistically using analysis-of-covariance techniques, but to the extent that unmeasured group differences are also likely, residual selection bias remains and the results are invariably equivocal. For example, in the Federal Trade Commission (FTC) study (1979), the coached group differed dramatically from the uncoached group on a number of indices related to SAT performance, such as having higher school grades and parental income, making it likely that they also differed on equally pertinent unmeasured factors such as motivation and parental education. Under these circumstances, the obtained score effects cannot be unequivocally attributed to the coaching treatment and must be interpreted as combined coaching/self-selection effects (Messick, 1980a; Stroud, 1980).

The standard prescription for avoiding selection bias is an experimental design with random assignment of participants to coaching treatment groups and noncoaching control groups, for only with random assignment can treatment effects be presumed to be independent of prior status on any of a host of personal or background characteristics. At the outset, no systematic differences are expected between treatment and control groups when they are constituted by random assignment, and if effective control conditions are instituted and maintained, the only systematic difference that should eventuate is that the treatment group will have been coached while the control group will not.

To date, three studies of SAT coaching have employed random assignment (Roberts & Oppenheim, 1966; Evans & Pike, 1973; Alderman & Powers, 1980) and, given the power of randomized experiments to obviate selection bias, one might ordinarily expect heavy reliance to be placed on their results. However, in all three instances the posttest used was a special administration of a retired form of the SAT or PSAT (Preliminary Scholastic Aptitude Test) rather than a regular SAT administration. In each case, for purposes of equity, the plan was to make coaching available to all study participants but to postpone access for the control groups until after the experimental posttest; the intent of administering a special SAT was to protect the control students from having test scores count on their college admission records before they had an opportunity to be coached. But precisely because they did not count, these special administrations may have been viewed to some degree as practice tests, thereby eliciting less motivation and effort than would a real SAT administration, especially for the uncoached control students.

Warning signals that this might have been the case were detected in two of the studies (Roberts & Oppenheim, 1966; Alderman & Powers, 1980), in which some of the control groups were found to exhibit score decreases in going from the pretest to the special SAT posttest when the expectation from national norm samples of test repeaters is for score increases over the same time period. Control-group score decreases were not observed in the third randomized study (Evans & Pike, 1978) or, for that matter, in all of the schools included in the other two investigations -- control-group scores in 3 out of 14 schools did not decrease in Roberts and

Oppenheim (1966) while 5 out of 8 did not in Alderman and Powers (1980). But there is no way of gauging how much the obtained score increases were influenced by the different motivational conditions of a special as opposed to a regular SAT.

Although the available studies of coaching for the SAT are methodologically flawed in one way or another, the various defects entail sufficiently divergent implications that any regularities or lawful consistencies across the studies would nonetheless be compelling. Most of the studies are subject to the influence of selection bias discussed earlier, which severely compromises interpretations as to the source or determinants of score effects -- in particular whether they may be unequivocally attributed to coaching experiences as opposed to personal or background characteristics of the (self-)selected students. In this regard, some of the studies involved control groups of uncoached students attending different schools from those of the coached students or else drawn from other extrinsic sources such as test-score files, thereby confounding coaching effects with school effects and numerous self-selection factors (Dyer, 1953; French, 1955; Dear, 1958; Federal Trade Commission, 1979; Stroud, 1980). In some other studies, control groups of uncoached students were specially constituted to match available samples of commercially coached students on a number of variables, but this still allows systematic differences between the groups on unmatched variables (Frankel, 1960; Whitla, 1962).

Another defect common to several of the studies is an unfortunate reliance on small samples of coached students, which results in imprecise estimates of score effects and a reduced likelihood that real effects will be detected as statistically significant (Frankel, 1960; Whitla, 1962; Coffman & Parry, 1967; Alderman & Powers, 1980). Moreover, as we have seen, some studies were subject to the unrealistic motivational conditions of a special as opposed to a regular SAT administration, which very likely introduced biases in the estimated sizes of score effects (Roberts & Oppenheim, 1966; Evans & Pike, 1973; Alderman & Powers, 1980). Finally, and by far the most troublesome from the standpoint of estimating and interpreting coaching effects, some of the studies had no control groups at all (Pallone, 1961; Marron, 1965; Coffman & Parry, 1967).

All of these studies are arrayed in Table 1, which lists for each coaching program the type of control-group design employed, the size and statistical significance of the obtained score effect, the amount of student contact time entailed, and the size of experimental and control samples. This information is given separately for the the Verbal (V) and Math (M) parts of the SAT. The studies are grouped into three sections in order to obtain separate estimates of average score effects from those studies having control groups, those without control groups, and those manifesting some evidence of distorted or biased results. Included in this latter group are those coaching programs which exhibited control-group score decreases (Roberts & Oppenheim, 1966; Alderman & Powers, 1980) as well as the study by Coffman and Parry (1967), which was marred by treatment-group score decreases and other indications of motivational difficulties. The estimates of score effect and of student contact time in Table 1 are as given in Messick and Jungeblut (1981), in which each of the studies is summarized in detail. The score effects for the control-group studies are either weighted intercept differences between experimental and control regression lines, when these were available from analyses of covariance, or else they are weighted average score increases of experimental over control groups, weighted in the case of multiple experimental or control groups by their respective sample sizes. The score effects for the uncontrolled studies were estimated by reducing the reported score gains by an amount approximating the corresponding experiential growth expected on the basis of both national norms

Table 1

SAT Score Effects in Relation to Student Contact Time in Coaching Studies With and Without Control Groups

Verbal

	Study	Control Group Design	Score Effect	Student Contact Time in Hours	N Exp/Control
Controlled	Dyer (1953)	Different School	4.6*	15	225/193
	French (1955) V & H	Different School	18.3**	9.1	161/158
	French (1955) Vocabulary	Different School	5.0*	4.5	110/158
	Dear (1958)	Same and Different Schools	-2.5ns	6	60/526
	Frankel (1960)	Same School Statistically Matched	8.4ns	15	45/45
	Whitla (1962)	Statistically Matched	11.0ns	5	52/52
	FTC (1979, Stroud, 1980) School A	Test-Score Files	31.7**	20	393/1729
	School B	Test-Score Files	3.2ns	12	163/1729
	Alderman & Powers (1980)				
	School C	Randomized	-2.75ns	10.5	22/17
	School E	Randomized	-.64ns	6	25/74
	School F	Randomized	.63ns	5	37/35
	School G	Randomized	10.00ns	11	24/70
	School H	Randomized	12.10ns	45	16/19
			14.4 weighted average (13 studies)		
Uncontrolled	Pallone (1961)				
	Short	None	41	45	20+
	Long	None	85(60)[a]	100[b]	80+
	Marron (1965)				
	Group 1	None	56	300	83
	Group 2	None	33	300	608
	Group 3	None	36	300	5
	Group 4	None	13	300	26
			39.0(36.5)[a] weighted average (6 studies)		
	RANK-ORDER CORRELATION .77 (19 studies)				
Suspect	Roberts & Oppenheim (1966)	Randomized	14.6*	3.0	154/111
	Alderman & Powers (1980)				
	School A	Randomized	20.30ns	7	28/22
	School B	Randomized	7.00ns	10	39/40
	School D	Randomized	6.82ns	10	40/43
	Coffman & Parry (1967)	None	4	48	19
			17.9 Weighted average (5 studies)		
	RANK-ORDER CORRELATION .62 (24 studies)				

Math

	Study	Control Group Design	Score Effect	Student Contact Time in Hours	N Exp/Control
Controlled	Dyer (1953)	Different School	12.9**	8.3	225/191
	French (1955)	Different School	11.0**	8.3	161/260
	Dear (1958) Long	Same and Different Schools	23.4**	15	71/116
	Dear (1958) Short	Same and Different Schools	21.5**	6	60/526
	Frankel (1960)	Same School Statistically Matched	9.6ns	15	45/45
	Whitla (1962)	Statistically Matched	-5.3ns	5	50/50
	FTC (1979, Stroud, 1980) School A	Test-Score Files	24.9**	20	393/1729
	School B	Test-Score Files	7.3ns	12	163/1729
	Evans & Pike (1973)				
	Group QC	Randomized	11.0*	21	145/129
	Group DS	Randomized	19.0**	21	72/129
	Group RM	Randomized	25.0**	21	75/129
			16.2 weighted average (11 studies)		
Uncontrolled	Marron (1965)				
	Group 1[c]	None	39	100	232
	Group 2[c]	None	53	100	603
	Group 3[c]	None	44	300	78
			30.2 weighted average (3 studies)		
	RANK-ORDER CORRELATION .71 (14 studies)				
Suspect	Roberts & Oppenheim (1966)	Randomized	8.1ns	3.0	188/122
	RANK-ORDER CORRELATION .74 (15 studies)				

[a] Because of a 25-point discrepancy between the values reported in Pallone's (1961) text and tables, it is not clear which score effect is correct. Both are reported here along with the corresponding weighted averages.

[b] Since students in this coaching course were also attending preparatory school full time, the amount of Verbal contact time might be more appropriately estimated as 300 hours (Messick & Jungeblut, 1981).

* Significance level < .05 ** Significance level < .01 ns Nonsignificant

data and control-group growth in similar studies. See Messick and Jungeblut (1981) for details.

To begin with, we note that a simple tallying of statistically significant and nonsignificant score effects is not very illuminating. Of the 17 Verbal score effects derived from control-group studies (including the suspect ones), 5 are statistically significant at the .05 or .01 level and 12 are not, but 10 of the latter values are based on treatment-group sizes of about 50 cases or less compared with more than 100 cases for each of the significant effects. If the 8 schools from the Alderman and Powers (1980) study are combined -- including the suspect values from Schools A, B, and D, which serve to inflate the composite estimate -- the resulting weighted pooled effect of 8.4 could be adjudged statistically significant at the .05 level, and the overall SAT-Verbal tally would become 6 significant and 4 nonsignificant. Of the 12 Math score effects from control-group studies, 8 are statistically significant (all but one at the .01 level) and 4 are not significant, two of the latter being based on coached samples of 50 or fewer cases. Although such tallies of significant and nonsignificant findings may tell us more about the size of samples and about the power of statistical tests than about the effectiveness of coaching, even this gross comparison tends to suggest that SAT-Verbal may be somewhat less responsive to coaching interventions than SAT-Math, which should not surprise us given the greater curriculum relatedness of SAT-M.

Inquiring about the average sizes of obtained score effects is not very helpful either, because one gets strikingly different answers from those studies having some form of control group versus those studies having no control group at all. Excluding the suspect studies and weighting in each case by the size of the associated coached sample, the weighted average SAT-V effect from control-group studies is 14.4 and that of SAT-M is 16.2, whereas the corresponding values from the uncontrolled studies are 39.0 -- or 36.6 (see Table 1) -- and 54.2. These values represent average numbers of SAT score points on a score scale ranging from 200 to 800 points, on which typical score distributions exhibit standard deviations of about 100 points.

Since the uncontrolled studies yield estimated score effects that are far out of line with those of the control-group studies, we must carefully appraise the credibility of the adjustments made for experiential growth in lieu of control-group comparisons, for at first glance these uncontrolled studies appear to warrant a more general discounting on methodological grounds alone. We must also consider the possibility that even though the score effects from the uncontrolled studies are out of line with those from the controlled studies, they may be in line with expectations derived from systematic program differences between the two types of studies -- in particular, the fact that the coaching programs in the control-group studies, with one exception, entailed about 20 hours of student contact time or less, whereas the programs in the uncontrolled studies, again with one exception, entailed about 300 hours of student contact time. However, before addressing these two questions in detail -- that is, how defensible are the estimated score effects from the uncontrolled studies and how credible are they as student outcomes from long-term coaching programs? -- we should first explore a further complication: In some cases, the program score effects reported in Table 1 obscure emergent interactions between the size of individual score effects and the personal or background characteristics of the students.

Although statistically significant interactions have been observed in a number of coaching studies, the findings do not readily cumulate into a coherent picture. For example, in a study of two highly selective independent schools for boys, Dyer (1953) found that for students not taking mathematics as seniors, those who were coached gained over 29 points more on SAT-M than those who were not coached; whereas for students taking mathematics at the time, those who were coached gained only 3 points more than those who were not coached. In contrast, although French (1955) noted the same pattern in attenuated form for boys, he obtained the reverse pattern for girls: Coached girls currently studying math exhibited score increases over the noncoached girls in two control schools of about 20 and 30 points, respectively, whereas the coached girls not taking math at the time exceeded their noncoached counterparts in the two control schools by only 1 and 4 points. Thus, an interaction between the size of SAT-M score effects and current enrollment in math courses was replicated for males but the reverse was found for females, suggesting that the relationship is either haphazard or systematically moderated by the student's sex. In opposition to Dyer's (1953) results, Dear (1958) found that the greatest coaching gains on SAT-M were for students currently studying math, but no comparison can be made with French's (1955) finding of a sex-moderated interaction because Dear's sample was not broken down for separate analyses by sex of student.

In regard to SAT-Verbal score effects, the only interactions observed to date were uncovered in Stroud's (1980) reanalysis of data from the FTC (1979) study. At one of the two coaching schools in the FTC study (School A in Table 1), students exhibited combined coaching/self-selection effects that were statistically significant for both SAT-V and SAT-M, but no interactions were discovered for any of the dozen or so variables that served as covariates in the analyses of covariance. However, at the other coaching school (School B), though nonsignificant score effects were obtained overall, two statistically significant and independent interactive effects on SAT-V were observed, one for race and the other for self-reported parental income. On the average, even though their number was quite small (N = 13), black students at School B exhibited significantly larger coaching/self-selection effects on SAT-V than nonblacks. In addition, students reporting low family income exhibited significantly larger coaching/self-selection effects on SAT-V than those reporting high family income (Messick, 1980a, pp. 46-51). In contrast to previous findings (Dyer, 1953; French, 1955; Dear, 1958), no differential score effects were uncovered for SAT-M, either as a function of years of math taken or of sex or of any of the other covariates.

This sporadic emergence of statistically significant interactions indicates that certain types of students, such as those with low family income or males not currently studying math, might exhibit larger than average score increases in some coaching programs. These are particular instances of a more general possibility that we should now be ever alert to -- that coaching programs, like other forms of teaching, may display differential effects for different types or groups of students.

Since particular types of students may exhibit larger than average score effects in some coaching programs and since the score effects exhibited in the uncontrolled studies were larger than the average effects obtained in the control-group studies, the discrepancy in the average size of effects in the two kinds of studies may be partly attributable to student differences as well as to program differences and design differences. This brings us back to the two questions that were temporarily left in abeyance -- how defensible are the corrections for experiential growth used in estimating score effects in the uncontrolled studies and how plausible are those adjusted score effects as student gains in ability due to coaching? The interpretive difficulties introduced by the complete absence of control groups were indeed compounded by the special and self-selected nature of the students in the uncontrolled studies. These studies (Pallone, 1961; Marron, 1965) were each conducted in private preparatory schools that specialized in providing high school graduates with a year of post-high school study aimed at securing admission to the U. S. service academies and selective colleges (the school involved in Pallone's investigation was one of the ten studied by Marron).

In the absence of control groups of similar students, it is difficult to appraise the import of the obtained score gains. Four methods of adjusting for experiential growth in ability over the pretest-posttest interval have been suggested as a means of salvaging these studies, but these proposals are all questionable for one reason or another. The authors of the original articles both suggested that the obtained score increases be compared with normal expectations of SAT gains typical of males during the senior year of secondary school, as revealed by College Board normative data (Pallone, 1961; Marron, 1965). This is not a very satisfactory comparison, however, because the preparatory school students in question, most of whom were high school graduates, were clearly not representative of high school seniors who take the SAT. Another suggestion was to compare the obtained results with average gains in national SAT administrations of junior- to senior-year retesters having the same initial average score levels as the preparatory school students (Slack & Porter, 1980). Again, this is not a very satisfactory comparison because these private school students were not a representative sample of the national population of test repeaters. Other suggestions involved a comparison of the original score increases with the average gains of control students in superior schools from other studies of proprietary programs (Pike, 1978) or with score gains of control students in other coaching studies who had average initial score levels roughly comparable to Pallone's and Marron's groups (Messick & Jungeblut, 1981).

The point is that in the absence of comparable control groups no generally satisfactory estimate of coaching or instructional effects can be made. But the fact is that these four proposed adjustments, though based on different and debatable rationales, yield corrections for experiential growth which are not radically different from one another. The lowest estimates of experiential growth were generally produced by the Slack and Porter (1980) procedure and the highest by Pike's (1978); on the average, the former fell about 10 points below the mean of all four adjustments and the latter fell about 8 points above the mean. The correction applied to each uncontrolled study in Table 1 was the average of all four proposed procedures (Messick & Jungeblut, 1981). This resulted in weighted average estimates of experiential growth in Pallone's case of 41 SAT-V points over a 12-month testing interval and, in Marron's case, of 23 SAT-V points and 25 SAT-M points over a 6-month testing interval. Given the highly self-selected nature of the preparatory school students and the fact that these gross procedures leave important factors of differential motivation and growth uncontrolled, these provisional values for experiential growth may still be somewhat underestimated -- and the corresponding adjusted score effects somewhat overestimated -- but they do not seem to be unreasonable. Indeed, they appear to be in the right ballpark.

The resulting weighted average score effects from the uncontrolled studies were 39 points for SAT-Verbal and 54 points for SAT-Math, which brings us to the question of how credible it is to find such large score effects associated with SAT coaching programs. Of course, with such highly self-selected students -- and especially in the absence of control groups or refined covariance adjustments -- these score effects represent combined coaching/self-selection effects, and their size may be as much a consequence of student characteristics as of program effectiveness. On the other hand, the preparatory school programs in question were not typical SAT coaching efforts, but rather entailed "full-time exposure to course content that is directly related to the verbal and mathematics College Board tests (both aptitude and achievement)" (Marron, 1965, p. 1), undertaken "for students in their final year of precollege work, including a large number of high school graduates who were completing a year of post-high school study" (Pallone, 1961, p. 655). It should come as no surprise if such a long-term instructional program with curriculum emphases on test-relevant knowledge and skill development were found to yield larger SAT score gains than short-term programs emphasizing test familiarization and practice, as is the situation with some of the control-group studies in Table 1. If this is the case, then the search for categorical answers to the simplistic question of whether or not coaching works is seen to be futile, for the computation of an overall average score effect across the range of studies in Table 1 would produce a misleading middle ground that fails to characterize the results of either the long-term or the short-term programs very well.

As a consequence, there is little recourse but to seek relational answers to more complex questions of degree. Instead of appraising the average size of score effects associated with coaching in general, we should rather appraise the size of particular score effects in relation to the amount and kind of coaching, asking how much student time and effort devoted to what kinds of coaching experiences yield what level of score improvements. In this approach, a natural first step is to correlate the size of obtained score effects with available indices of program characteristics. Unfortunately, data on program characteristics that are sufficiently commensurable across studies to permit a detailed comparison and evaluation of coaching procedures is typically not available. Indeed, no coaching study as yet has systematically investigated the kinds of instructional methods and materials that may be most effective in improving SAT performance. Although program descriptors that apply uniformly across studies are rare in the literature on coaching for the SAT, the amount of student contact time in each coaching program does provide a generally applicable index of student effort. Since, for the relevant research studies, those coaching programs involving relatively high amounts of student contact time also entailed structured curricula emphasizing knowledge and skill development while the relatively low-contact programs emphasized test review and practice, this index embodies a confounding of program characteristics in these data. Thus, student contact time in this context can also be viewed as a proxy for increasing curriculum structure and increasing emphasis on skill development.

When all of the studies in Table 1 are ranked in order of their obtained score effects and this ranking is compared with their ordering in terms of student contact time in the respective coaching programs, the two rank orders prove to be remarkably similar. As seen in Table 1, the Spearman rank-order correlation coefficient between SAT-V score effect and student contact time across all 24 Verbal coaching studies is .62, while that for SAT-M across all 15 Math coaching studies is .74. Both coefficients are statistically significant at the .01 level. If the five suspect studies are deleted from the calculation for SAT-V, the new correlation is .77 across 19 studies, which is also significant at the .01 level. Thus, as anticipated, the suspect studies add considerable noise to the data set, but the relationship between SAT-V score effect and student contact time remains strong with or without them. The results for SAT-M, in contrast, are hardly affected by the lone suspect Math coaching study: Upon its deletion, the new correlation is .71 across 14 Math studies, which again is significant at the .01 level.

In interpreting these sizable correlation coefficients, it must be remembered that as rank-order statistics they imply a regular monotonic relationship between student contact time and score effect but not necessarily a linear one. Indeed, even though the time dimension is covered in only a fragmentary fashion by the available studies, when the magnitude of score effect is plotted against student contact time, the relationship is distinctly nonlinear. To demonstrate this, when linear regression lines were fit to score effects from coaching programs requiring less than 50 hours of student contact time (all of which included control groups), the extrapolated predicted score effects for coaching programs entailing 300 hours of contact time (none of which included control groups) were much higher than the score effects actually obtained. In the case of SAT-V, the predicted 300-hour score effects consistent with linearity were two to three times as large as the obtained average score effects, whether actual or adjusted, and for SAT-M, three to four times as large. This marked linear overprediction of score effects for higher values of student contact time clearly indicates a phenomenon of diminishing returns in coaching effectiveness (Messick & Jungeblut, 1981).

Diminishing Returns in SAT Coaching Effects

As is frequently the case with diminishing returns, a logarithmic transformation of the time dimension provided a much better representation of the functional relationships. To illustrate, when linear equations were fit to the regression of score effect on log contact time for those coaching programs requiring less than 50 hours, the extrapolated predicted values for 300-hour programs (33 points for Verbal and 44 points for Math) now deviated from the corresponding obtained average adjusted score effects by only about 3 to 5 points for SAT-V, depending on which of Pallone's (1961) discrepant values are included, and by 10 points for SAT-M. Regression lines fit to all of the data in log time were very similar to those regression lines based only on control-group studies, differing from each other by only 2 or 3 points for SAT-V and 8 points for SAT-M at the benchmark of 300 hours (Messick & Jungeblut, 1981). These logarithmic equations based on the full range of data (excluding the suspect studies and Pallone's discrepant values) are plotted in Figure 1, transformed so as to relate SAT score effect to student contact time in real rather than log hours.

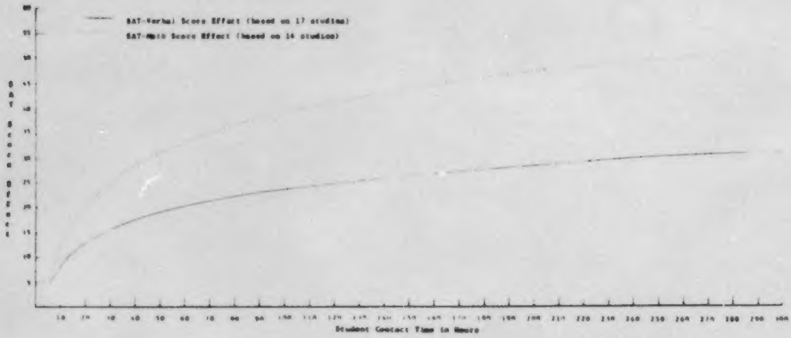


Figure 1

Expected SAT Score Effects Associated with Given Hours of Student Contact Time Separately for Verbal and Math According to the Logarithmic Models Fit to Available Coaching Data (i.e., 20 hours of Verbal coaching yields an expected score effect of about 13 SAT-V points, and, another 20 hours of math coaching yields an expected SAT-M score effect of about 21 points)

As can be seen from Figure 1, according to the logarithmic models fit to the obtained score effects from the available SAT coaching studies, a Verbal coaching program involving about 10 hours of student contact time would be expected to be associated with average SAT-V score effects of 8 or 9 points, while a 10-hour Math coaching program would be associated with average SAT-M effects of about 12 or 13 points. Similarly, the corresponding values for a 20-hour Verbal and a 20-hour Math program are 13 SAT-V points and 21 SAT-M points; for 30-hour programs apiece, 16 SAT-V points and 25 SAT-M points; for 50 hours each, 19 and 31 points; for 100 hours each, 24 and 39 points; for 200 hours each, 28 and 47 points; and, for 300 hours each, 31 and 52 points. Or for uniform 10-point increases in expected average score effect, the associated hours of contact time would be as follows: About 12 hours of Verbal coaching or 8 hours of Math coaching for an average of 10 SAT points; about 57 hours of Verbal or 19 hours of Math coaching for an average of 20 score points; about 260 hours of Verbal or 45 hours of Math for an average of 30 points; and, about 1185 hours of Verbal or 107 hours of Math for an average of 40 points. Thus, arithmetically increasing amounts of score effect are associated with geometrically increasing amounts of student contact time. For uniform 10-point increases in average score effect, as in the illustration just presented, the multiplicative constant carrying one amount of student contact time into the next higher amount is about 4.57 for SAT-V and about 2.37 for SAT-M. The fact that Verbal entails much more student contact time than Math for the same amount of average score effect -- at least for these logarithmic models fit to available SAT coaching data -- is consistent once again with the expectation that Math, being more curriculum related than Verbal, should be relatively more responsive to coaching intervention.

It must be emphasized that these logarithmic functions are based on existing data from available studies, all of which involved students who to a large degree were motivated to increase their test scores through coaching. Since such motivated students are likely to have been highly task oriented, it is not unreasonable under these circumstances that student contact time should be found to be directly related to average test score increases (Stallings, 1980). But we must remember two cautionary notes: One, that increases in student contact time are confounded in these studies with increasing curriculum emphases on knowledge and skill development, so that other program characteristics that contact time may be a proxy for should also be taken into account in interpreting the relationships and determining expectations. The other derives from the pervasiveness of methodological flaws in the extant studies, which compels us to consider the possible role of selection bias in interpreting the basis of the functional relationships -- many of the longer-term programs, such as the preparatory schools in Marron's (1965) study and commercial coaching school A in the FTC (1979) study, were not only associated with larger score effects but were also highly subject to self-selection bias. With these caveats in mind, it appears likely that improvement of the comprehension and reasoning skills measured by the SAT, when it occurs, is a function of the time and effort expended and that each additional score increase may require geometrically increasing amounts of time and effort. This developmental pattern of gradual and diminishing increases in response to coaching or instructional experiences is just about what one might expect with measures of stable, though developing abilities.

Controversies Over Implications

Before we can properly appraise the import of an apparently curvilinear relationship between student contact time and score effects associated with coaching, we must carefully consider the different ways in which coaching might operate to improve test scores. With respect to implications for educational and testing practice, it matters what processes underlie improved test performance -- in particular, whether any increased test scores attributable to coaching reflect stable improvements in the verbal and mathematical reasoning abilities measured by the SAT as opposed to improved facility in overcoming inadvertant sources of test difficulty unrelated to these reasoning skills (such as test anxiety or unfamiliarity with different item formats and test-taking strategies), or some combination of the two. We must also consider the extent to which any obtained score increases associated with coaching are practically worthwhile in relation to the investment of student time, effort, and financial resources required and in light of multitudinous factors that complicate this judgment. Depending on their source and practical utility, SAT score gains attributable to coaching -- if they were to occur with any regularity -- would have important policy implications for either educational practice or testing practice, or both.

Implications for Student Performance and Test Validity

Of the potential ways in which coaching may function to improve test scores, three major possibilities can be distinguished (Messick, 1980a): First, some coaching programs may genuinely improve the abilities and skills measured by the test, resulting in commensurate increases in test scores. Such score increases reflective of improved abilities should threaten neither the construct validity nor the predictive utility of the SAT, for any lasting improvements in verbal and quantitative reasoning abilities should lead to score increases on tests measuring these constructs and should also serve the student well in criterion situations entailing these abilities, as is the case in school and college learning. Second, some coaching programs may enhance test-taking sophistication or reduce the anxiety often associated with taking tests, resulting in increased test scores that are now more accurate assessments of student ability. Such score increases reflective of improved test wiseness, by virtue of revising previous scores that were inaccurately low because of construct-irrelevant test difficulty, should lead not only to more accurate assessment of abilities but also to enhanced predictive validity of the test. Third, some coaching programs may teach test-taking stratagems and answer-selection tricks, resulting in increased test scores that are inaccurately high as assessments of student ability. Such score increases reflective of acquired artifice should not only dilute the construct validity of the test but jeopardize its predictive validity as well. Some coaching programs, of course, may produce none of these effects or more than one in various combinations, so that in any particular instance some mix of improved abilities, improved test wiseness, and improved artifice might contribute to improved test performance.

The first two possible outcomes, if they were realized, would be good both from the standpoint of student performance and from the standpoint of test validity. The third possibility is probably only a minor problem with well-constructed tests, because professional test makers strive to minimize the use of complicated or tricky item formats and to avoid items that may be answered on the basis of clues unrelated to the abilities tested. Moreover, test makers also strive to attenuate the import of the second possible outcome of coaching by providing test-familiarization materials and practice tests to all candidates as well as advice on guessing, pacing, reviewing, and the like. "With well-made tests the concern is not that students will learn how to capitalize on extraneous clues to inflate their test scores, for such clues should be rare and are further reducible through test analysis, but rather that students learn to cope most advantageously with the test as a standardized vehicle for demonstrating their abilities" (Messick, 1980a, p. 65).

Thus, of three main possible outcomes of coaching, two are good for both student performance and test validity while the third is a minor and rarely demonstrable problem with professionally developed tests. The second possibility -- that improved test wiseness might yield increased test scores of greater accuracy -- although potentially critical for certain types of students who are relatively unpracticed and unoriented in the ways of standardized testing, is of debatable importance for the generality of students. The principal reason for this surmise is that various coaching programs have attempted to enhance test wiseness by drilling the student in different approaches to different item formats and to allay anxiety by providing feedback on effective item performance, but the available fragmentary research evidence indicates that if this is all that is done such coaching will have little impact on SAT scores (Messick, 1980a). Coaching programs that emphasize test familiarization and practice appear to be associated with small score effects, if any, whereas programs that include skill-development components tend to be associated with larger score effects which, on the whole, seem to occur more for Math than for Verbal. This can be seen in Table 1 once we note that the former programs tend to be relatively short-term while the latter programs are relatively long-term.

This leaves us with the first possible outcome of coaching -- namely, that effective coaching programs genuinely improve the verbal and mathematical reasoning abilities measured by the SAT -- and with the task of ascertaining the credibility of such an outcome given the paucity of directly relevant research evidence. No study of coaching to date has systematically addressed this issue of improved abilities. Rather, all of them have focussed on the prior issue of first demonstrating significant score increases associated with coaching before inquiring into their causes. Since this second step has not yet been taken, at this point we can only conjecture as to the likelihood that obtained score increases reflect improved reasoning skills.

An important related question pertains not only to the improvement of abilities through coaching but to the stability of that improvement over the long term. If effective coaching does improve abilities, this might occur via the development of new skills or, more likely, by the strengthening, honing, and refining of existing skills through exercise and challenge. Although increased test scores attributable to coaching may reflect real ability improvement at the time of testing, this improvement may be relatively enduring or transitory depending on whether critical levels of exercise and challenge are maintained in the student's learning environment until the ability gains become stably consolidated.

Since coached scores, if they indeed represent improvements in ability, should be just as valid or more valid than uncoached scores in predicting academic success, at least in the first year of college, one might turn to such predictive data if it existed for evidence of improved abilities. However, such an approach is both indirect and ineffectual. It is indirect because it requires not only that reasoning abilities be improved by coaching but that the improved skills be both durably maintained over the freshman year and effectively utilized in college performance, whereas numerous countervailing factors -- such as social adjustment problems and heavy involvement in athletics or other extracurricular activities -- may contribute to scholastic underachievement and the gradual erosion of academic skills. It is ineffectual with respect to inferences about improved abilities because the validity of coached scores might derive not from improved abilities but from improved accuracy in the assessment of existing abilities attained through enhanced test wiseness. Thus, high predictiveness would provide positive evidence that increased test scores attributable to coaching reflect either stable improvements in abilities or more accurately assessed abilities which in either case were well utilized in college performance, but low predictiveness would constitute negative evidence only if plausible alternative explanations for poor academic performance could be discounted.

Despite these interpretive pitfalls, one coaching study -- that undertaken by Marron (1965) at ten specialized preparatory schools -- tentatively explored this indirect route by appraising the extent to which SAT scores after long-term coaching predicted freshman class standings at the U. S. service academies and selective colleges. As a further complication, however, relatively crude and approximate methods of equating were required in Marron's study to achieve some semblance of distributional comparability across the service academies and across the widely diverse colleges that the preparatory students dispersed to. Although the outcome is thereby rendered admittedly tenuous, the findings nonetheless suggest that at the service academies the students did less well academically than the test scores predicted, whereas at the selective colleges the distributions of class standings and test scores did "not seem to be inconsistent" (Marron, 1965, p. 22).

Thus, the possibility that effective coaching may genuinely improve the comprehension and reasoning skills measured by the SAT is neither strongly supported nor strongly countered by available data, mostly because so little of it is directly pertinent to the issue. What would be more germane are studies relating both uncoached SAT scores and SAT scores after various kinds and amounts of coaching to multiple measures of developed abilities and cognitive processes, along with indices of test-taking sophistication and test anxiety. One might thereby determine the extent to which coached and uncoached scores reflected the same cognitive ability factors, the extent to which they entailed format-specific variance unrelated to other methods of measuring these abilities, and the degree to which they were differentially affected by test-taking skills and anxiety levels. But in the absence of such studies, we have little recourse but to glean whatever clues we can from available analyses.

To assess even tentatively which way the existing data point, we must rely on indirect signals and bits of circumstantial evidence, most of which are discernible in Figure 1. After all, what research evidence there is about coaching for the SAT has been summarized in Table 1 and statistically integrated into the curves of Figure 1. Those curves suggest that SAT score effects -- as embodied in a logarithmic model fitting the available fragmentary data -- are still increasing, minutely but steadily, up to 300 hours of student contact time and beyond. Each additional amount of score increase entails geometrically more student contact time, which is a pattern more suggestive of developing abilities than of either enhanced test wiseness or acquisition of subject-matter knowledge.

Subject-matter achievement ought to increase at a faster or less diminishing rate in response to tuition, as indeed it did in Marron's (1965) study, in which weighted average score gains on achievement tests of English composition were almost one and a half times as large as unadjusted weighted average gains on SAT-V while gains on intermediate and advanced Math tests were roughtly one and three-quarter times as large as on SAT-M. In contrast, it is most unlikely that enhanced test wiseness or anxiety reduction would still be accruing positive effects, however diminishing, after 300 hours or even 100 hours. Given the circumscribed nature of the information and advice to be imparted and the inevitable redundancy of repeated test practice per se, coupled with the modest results of coaching programs emphasizing test familiarization and item review, one might expect the effects of test-wiseness training alone to peak and level off within 10 or 20 hours of student contact time or less. After all, 10 to 20 hours of test-wiseness training would be comparable to a one-credit college course in test taking. Consistent with this conjecture, the faster initial growth rates in the curves of Figure 1 indeed suggest that score effects associated with the first 10 to 20 hours or so of student contact time may possibly reflect a combination of enhanced test wiseness and developing abilities, with the former becoming progressively less important and the latter progressively more important in the attainment of larger and larger score effects.

Implications for Guidance and Admissions

SAT score increases attributable to coaching, if they occur with any certitude, may have important implications not only for testing practice but also for educational practice -- for learning and instruction as well as for guidance and admissions. Before we can ponder potential implications, however, we must confront the question of practical utility -- how large does an obtained coaching effect need to be to have practical significance for various purposes? As we shall see, since a number of factors must be taken into account simultaneously, this is a complicated question which can only be answered conditionally. Nonetheless, it would help to clarify if not resolve the issue if we could underscore some of the relevant factors while at the same time discounting other ostensible answers that turn out to be largely irrelevant to the utility of coaching effects.

In this regard, one commonly mentioned benchmark is the standard error of measurement of the test, which typically hovers around 30 score points or so for both SAT-V and SAT-M. The standard error of measurement is the standard deviation of measurement errors from whatever source -- of major concern in connection with a multiple-form test such as the SAT are those errors incumbent upon taking one test form as opposed to other parallel or equivalent forms. A range of plus or minus one or two standard errors of measurement around a student's observed score is usually taken as a confidence interval serving to bound, with a given probability, the student's true score. Compared with this band of plus or minus 30 to 60 points of expected error, most of the score effects associated with coaching programs in Table 1 appear to be relatively small. But these score effects, by and large, represent differences in average score increases of coached groups over control groups, which automatically takes the standard error of measurement into account since it operates in similar fashion in both the coached and uncoached samples.

From this perspective, the standard error of measurement is irrelevant to the appraisal of average score effects associated with coaching, because these effects reflect differences over and above the influence of measurement error. It is also largely irrelevant to the appraisal of individual score effects attributable to coaching, provided they are calculated as residual scores around regression lines based on uncoached or pooled samples, as in analysis of covariance (Cochran, 1968). Such residual scores also estimate coaching effects over and above the standard error of measurement, in addition to being adjusted for contaminating variance on any covariates included in the regression equation -- although the size of the standard error does affect the size of the individual residuals. However, the standard error of measurement is by no means irrelevant to individual score gains calculated as simple differences between two testings before and after coaching. In this case, negative errors of measurement in the first testing coupled with positive errors in the second would inflate any coaching effect, whereas positive measurement errors in the first testing coupled with negative errors in the second would attenuate it.

Another reason why the standard error of measurement is not an apposite yardstick for gauging average coaching effects is that measurement error is a random process affecting all test takers irrespectively, whereas coaching is a deliberate and directed process selectively affecting only those test takers who engage in it. There is no question of unfairness in connection with error of measurement, for it constitutes the luck of the draw in how well the sample of items on a particular SAT form happens to match the individual's functional capabilities on a given day. That on a different day or a different test form the individual's score might be somewhat higher or somewhat lower is what is meant by error of measurement, and it applies to everyone on a randomly probabilistic basis. In contrast, demonstrably effective coaching, being systematic and selective rather than random and universal, would inevitably raise questions of unfairness; as we shall see when such questions are addressed in detail in the next section, however, they are neither new nor unique to coaching. Nonetheless, in terms of social implications, since a relatively large standard error universally applied might be tolerable while a same-sized coaching effect selectively applied might not, there are ethical as well as logical and statistical reasons to avoid comparing the two.

Other oft invoked contexts for evaluating the size of coaching effects are the SAT score scale, which ranges from 200 to 800 points, and the number of additional items correct implicit in a given score effect. For example, a coaching effect of 20 SAT-V points corresponds to about three additional Verbal items correct while 20 SAT-M points corresponds to roughly two additional Math items correct. A 20-point coaching effect thus seems trivial when viewed in terms of additional items correct or in terms of a 600-point potential score range. But the choice of a score scale is arbitrary and, within the limits of adequate reliability and feasible testing time, so is the number of items in the test.

Although the 200- to 800-point SAT score scale is immaterial for gauging the size of coaching effects, the distribution of test scores along that scale is definitely not. In particular, the standard deviation of the SAT score distribution, which is normally in the neighborhood of 100 points for both SAT-V and SAT-M, provides a highly pertinent yardstick for assessing coaching impact. Although the numerical value of the standard deviation is tied to the arbitrary score scale, as is the metric of the coaching effects, what is not arbitrary is the number of standard deviation units that a given coaching effect represents. Thus, a 20-point coaching effect is roughly a fifth of a standard deviation of the score distribution, and it remains a fifth of a standard deviation under any linear transformation of the score scale.

By casting coaching effects in standard deviation units, we are in a position to translate score gains attributable to coaching into corresponding increases in percentile rank. That is, for normally distributed test scores in general, we can now determine the improvement in a student's percentile rank attendant upon a given score effect, but this percentile change is variable, being contingent upon initial score level prior to coaching. For example, if the initial uncoached score were two standard deviations below the mean, a 20-point score effect would yield a percentile improvement of 1.3; if one standard deviation below the mean, a 20-point effect yields an improvement of 5.3 percentiles; right at the mean, 7.9 percentiles; one standard deviation above the mean, 4.4 percentiles; and, two standard deviations above the mean, only .9 percentiles (see Table 2). Thus, the largest improvement in percentile rank occurs for initial scores at the mean of the score distribution, with the amount of improvement falling off as initial scores deviate from the mean in either direction. Furthermore, a given score effect yields slightly larger percentile improvements for initial uncoached scores below the mean in comparison with the improvements associated with corresponding scores equally far above the mean. Although the SAT score scale is not precisely normal and the standard deviation varies slightly from year to year, the corresponding percentiles and

Table 2

Percentile Gain Associated with 20- and 30-point Coaching Effects
for Initial Scores in Standard Deviation (SD) Units

Normal-Curve Values (SD = 100)				1979-1980 SAT Values for College-Bound Seniors					
				Verbal (SD = 110)			Math (SD = 117)		
		%ile Gain	%ile Gain		%ile Gain	%ile Gain		%ile Gain	%ile Gain
SD Above or Below Mean	Initial %ile	20-point (.25SD) Coaching Effect	30-point (.35SD) Coaching Effect	Initial %ile	20-point Coaching Effect	30-point Coaching Effect	Initial %ile	20-point Coaching Effect	30-point Coaching Effect
+2.0	97.7	.9	1.2	96.4	1.6	1.6	97.0	1.0	1.0
+1.5	93.3	2.2	3.1	91.9	2.0	3.0	91.3	2.8	3.9
+1.0	84.1	4.4	6.2	82.8	4.0	5.6	81.6	4.0	6.0
+ .5	69.2	6.7	9.7	68.7	5.1	8.9	66.8	6.1	8.6
0	50.0	7.9	11.8	50.2	7.0	10.0	49.8	6.4	9.0
- .5	30.9	7.4	11.2	31.6	5.2	9.9	33.0	5.3	8.3
-1.0	15.9	5.3	8.3	14.8	6.0	8.4	16.7	5.0	8.0
-1.5	6.7	3.0	4.8	5.9	2.9	4.0	5.1	3.1	5.0
-2.0	2.3	1.3	2.2	1.0	1.4	2.4	1.0	.2	1.0

percentile changes in the 1979-1980 national percentile tables for college-bound seniors are within a point or two of these normal-curve values, as seen in Table 2.

The percentage of test takers that a student surpasses by virtue of a specific coaching effect may be keyed to a particular reference group of interest, whether college-bound seniors or females or applicants to a specified college or whatever. This is accomplished by dividing the obtained score effect by the standard deviation of test scores for the corresponding reference group and then appraising the attendant percentile gain over initial score by means of normal-curve tables where appropriate -- or else by using local or targeted percentile tables.

With respect to the practical consequences of such percentile improvements, we should note that low-scoring students, who most need whatever benefits coaching might accrue, gain relatively little in percentile rank from a 20-point score effect. The same holds true for high-scoring students, who nonetheless might strongly desire any additional benefits derivable from coaching if they aspire to highly selective colleges. Assuming a 20-point coaching effect, a student with an initial uncoached score 1.5 standard deviations below the mean would rise three percentiles (from a percentile rank of 7 to a rank of 10), whereas a student initially scoring 1.5 standard deviations above the mean would gain about two percentiles (from a percentile rank of 93 to 95). For scores initially at the 50th percentile, a 20-point coaching effect would yield about an 8 percentile improvement.

Given current admissions practices, it is unlikely that percentile improvements of these magnitudes would substantially influence admissions decisions, except perhaps for students whose scores are at the margins of admissibility to variously selective institutions at whatever level. Even here, other factors are likely to outweigh a few additional percentiles on the SAT, unless rigid cutting scores are mandated. Nor is it likely that such improvements in percentile rank -- from 7 to 10, from 50 to 58, from 93 to 95 -- would dramatically alter student self-concepts and aspirations, although any improvements in performance or ability should be realistically capitalized upon in student guidance. Since popular usage seems to stress gross categories -- such as being in the top half, third, fifth, or tenth of the group -- any percentile changes that bridge those categories might have greater personal import psychologically, but there is little evidence that they have special practical significance over and above the percentile change itself. For example, a 20-point coaching effect would move a student initially scoring half a standard deviation below the mean from the bottom third to the middle third of the norm group, albeit from the top of the lower third to the bottom of the middle third, but it is doubtful that this gross characterization would outweigh in practical decisions the actual change in percentile rank from 31 to 38.

The outcome for a 30-point coaching effect is not radically different. As given in Table 2, the corresponding percentile changes are about 5 percentiles for initial scores 1.5 standard deviations below the mean (from 7 to 12), less than 12 percentiles for initial scores at the mean (from 50 to 62), and 3 percentiles for initial scores 1.5 standard deviations above the mean (93 to 96). In this case of a 30-point coaching effect, however, the percentile changes close to the mean, having exceeded a decile, are more noticeable and might have practical significance in selective instances. But coaching effects of these magnitudes, judging from the available research evidence, are not easily attainable, especially in Verbal (see Table 1). According to Figure 1, an expected 20-point SAT-V score effect would be associated with about 57 hours of contact time for motivated task-oriented students in a Verbal coaching program, while a 20-point SAT-M effect is associated with about 19 hours in a Math coaching program. An expected 30-point score effect would be associated with 260 hours for Verbal and 45 hours for Math. These score effects, of course, are average values expected to be associated with coaching programs of these durations, so some individuals would achieve variably larger gains with these amounts of contact time while other individuals would attain variably smaller gains. Moreover, given the possibility of interactions between score effects and various personal or background characteristics of the participants, certain groups or types of students might experience markedly higher or lower increases than that typified by the average score effect, but such interactions have not been well documented.

If one focusses not on the test as a predictor of college performance but on the predicted performance itself, another potentially useful yardstick for gauging coaching effects becomes apparent -- namely, the attendant gain in predicted criterion performance or college freshman grade-point average (GPA) associated with an obtained score effect. However, since this predicted criterion gain is dependent on particular grade criteria which vary widely from college to college and on the number and validity of other predictors included in the specific regression equation, it is woefully ad hoc as a yardstick for appraising coaching effects.

In order to get some inkling of the likely magnitude of such predicted criterion gains, however, let us consider some characteristic median correlations based on 827 SAT validity studies (Educational Testing Service, 1980). The median validity coefficient in predicting college freshman grade-point average (G) was .52 for high school grade-point average (H), .37 for SAT-V (V), .32 for SAT-M (M), and .58 for all three predictors combined. The corresponding multiple-regression equation in standard-score form is $\hat{Z}_G = .43Z_H + .20Z_V + .10Z_M$. If a student's test scores are augmented by a specific coaching effect in Verbal (Z_{VC}, in standard deviation units) and a possibly different coaching effect in Math (Z_{MC}), the resulting prediction is $\hat{Z}_{G_C} = .43Z_H + .20(Z_V + Z_{VC}) + .10(Z_M + Z_{MC})$. The gain in predicted criterion performance associated with these coaching effects is therefore $(\hat{Z}_{G_C} - \hat{Z}_G) = .20Z_{VC} + .10Z_{MC}$. Thus, for typical SAT standard deviations of about 100 points, a 20-point Verbal coaching effect yields an increment of .04 standard deviations in predicted freshman grade-point average while a 20-point Math coaching effect yields an increment of .02 standard deviations, resulting in a combined gain in predicted criterion performance of .06 standard deviations. For grade-point averages on a 0- to 4-point scale, the median standard deviation over a fifteen-year period was approximately .66 (Bejar & Blew, 1981). Therefore, the combined gain in predicted criterion performance associated with 20-point coaching effects in both Verbal and Math would amount to about .04 units of GPA.

In general, for linear regression, the gain in predicted grade-point average associated with coaching is determined by multiplying the obtained Verbal or Math coaching effect in standard deviation units by the respective beta coefficient for SAT-V or SAT-M (a number usually less than one). As a consequence, given standard deviation units of coaching effect will typically yield less than half that amount in standard deviation units of predicted criterion gain, often considerably less -- a single test predictor with a relatively high validity coefficient of .5 generally has its unique contribution even further reduced by the addition of other valid correlated predictors. Thus, even with relatively high test validities, a 30-point coaching effect (.3 standard deviations) would usually yield a gain of less than .15 standard deviations in predicted criterion performance. In terms of predicted grade-point average, then, in contrast to SAT percentile gains, an even smaller percentage of students would be surpassed by virtue of a particular coaching effect. Furthermore, to the extent that other valid predictors are employed in addition to the test and that other background factors influence admissions decisions, the import of any obtained coaching effect would be correspondingly muted in practice.

As has already been suggested, any practical utility that coaching effects might convey is likely to be greatest for that subgroup of students whose uncoached scores are at the margins of admissibility to selective institutions of their choice, in whatever score ranges those institutions deem marginal. But it should be remembered that score increases attributable to coaching may arise in three possible ways, and although any one of them might move a student from inadmissible to admissible status in terms of test scores, they have different implications for subsequent college performance. These three types of coaching effects are schematized in Figure 2 in relation to admissible and inadmissible ranges of test scores and to successful and unsuccessful ranges of criterion performance.

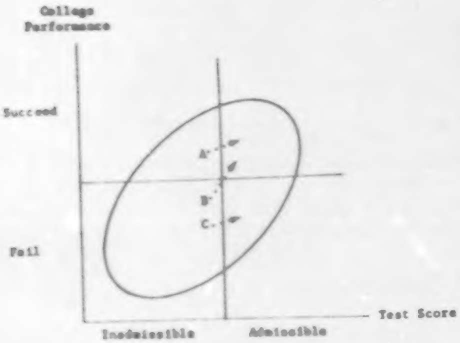


Figure 2

Three Possible Coaching Effects in Relation to College Performance (Adapted from Hammond, 1980)

Type A coaching effects in Figure 2 derive from enhanced test-taking sophistication or reduced test anxiety, by means of which uncoached scores that were inaccurately low because of construct-irrelevant test difficulty are raised to levels more accurately descriptive of student ability. Some students invalidly deemed inadmissible in terms of the formerly inaccurate scores might now be validly deemed admissible in terms of the more accurate assessment of ability levels. Type B coaching effects reflect genuine improvements in the comprehension and reasoning abilities measured by the SAT. Some students validly deemed inadmissible in terms of low prior ability levels might now be validly deemed admissible by virtue of improved abilities sufficient for successful college performance. Type C coaching effects derive from test-taking stratagems and answer-selection tricks, resulting in increased test scores that are inaccurately high as assessments of student ability. As a consequence, some students validly deemed inadmissible in terms of the more accurate prior scores might now be invalidly deemed admissible by virtue of inaccurately high scores masking insufficiently high ability levels.

Students, especially those from diverse cultural backgrounds, stress type A effects; commercial coaching schools claim type B effects; and, professional testing organizations denigrate type C effects (Hammond, 1980). As we have seen, there may be a modicum of substance in all three positions, but the important point is that these potential effects are neither equally likely nor equally consequential. For example, types A and B, if realized, should enhance both student performance and test validity while type C should be only incidental with professionally made tests. To the extent that they occur, type A has implications for testing practice and type B for educational practice, and both raise questions of equity of access to effective coaching programs, which is the next -- and final -- set of interrelated issues to be addressed.

Implications for Educational and Testing Practice

If coaching were to substantially increase test scores without correspondingly improving the abilities measured, there would be important implications for testing practice. Such an outcome would imply that the test or the testing experience entails unintended sources of difficulty that can be at least partially overcome by special preparation -- anxiety over being evaluated, for instance, or unfamiliarity with item formats or test-taking strategies. In addition, effective coaching might reduce the operative difficulty level of some test items by teaching answer-selection tricks. As we have seen, the latter possibility is unlikely for professionally made tests, but to the extent that it does occur, the affected items should be revised or eliminated and the procedures for test development and analysis tightened. Although the former possibility -- that enhanced test wiseness or anxiety reduction might lead to sizable score increases -- is also unlikely to be of major consequence for the generality of students, it may be of special significance for certain groups or types of students, such as test novices or those who are not culturally attuned to standardized testing. This suggests -- even though the available studies find only small average effects, if any, associated with coaching programs stressing test familiarization and practice -- that in the interest of equity all test candidates should have an opportunity to familiarize themselves with appropropriate test formats and to practice recommended test-taking strategies. Such familiarization might be accomplished, for example, through the diligent use of practice tests and advisory materials similar to those routinely distributed by the College Board. Issues of equity of access to such test-wiseness coaching programs and materials become important mainly to the extent that individual differences in test-taking skills per se influence test scores.

On the other hand, if score increases attributable to coaching represent commensurate improvements in the verbal and mathematical abilities measured by the SAT, then there would be important implications for educational practice and social policy. To begin with, the functional characteristics of any SAT coaching programs that prove to be effective would have direct implications for instructional practice, for although there is little question that the comprehension and reasoning skills measured by the SAT are learned, there are large questions about how they can be taught. And if they can be taught, their status as explicit objectives of school learning must be addressed. Since comprehension and reasoning abilities facilitate educational accomplishments generally -- the SAT is related, after all, not only to college performance but also to high school performance -- school programs stressing ability development as well as subject-matter learning should have synergistic effects on current and subsequent educational achievement. In addition, as Cronbach and Snow (1977) insist, "the most important safeguard against rigidity of streaming is to make the development of aptitude a deliberate goal of instruction" (p. 521). All of this implies, assuming comprehension and reasoning skills are teachable, that the most efficacious approach to elementary and secondary education -- as well as the soundest long-range mode of preparation for the SAT -- would be school programs that integrate the development of thought with the development of knowledge.

If coaching techniques for fostering ability development could be specified which were not an integral part of the regular school experience, then the issue of equity of access to special coaching programs arises. The SAT measures the current level of developed scholastic abilities facilitative of academic learning. Whether this current level of developed abilities derives in part from special coaching programs pointed toward improved test performance or from general educational programs pointed toward improved school performance or from extensive experiential learning in nonschool settings -- that is, from coaching or instruction or experience -- is indistinguishable to the SAT. Thus, the issue of equity of access to coaching programs that are effective by virtue of ability development, if such could be identified, is similar to the issue of equity of access to effective school programs or effective life experiences, for they each have similar consequences for test scores. Such ability-enhancing coaching programs raise important equity issues of educational access, to be sure, but they are not new equity issues nor are they unique to coaching.

Epilogue

To say that coaching for scholastic aptitude tests such as the College Board SAT is controversial is certainly true, but it does not come close to conveying the depth or dimensions of disagreement -- as in any long-standing controversy, there are deep-seated value issues entailing multiple value judgments. Nor does it help, necessarily, to trace the arguments of various parties, for often as not they are speaking past one another by virtue of different assumptions about the meaning of aptitude or of coaching and about how to pose the basic question of coaching effectiveness. Moreover, even if there were agreement on the meaning of terms and on the nature of the research question, there would likely be disagreement over the adequacy of the evidence and its consequent implications. It does help, however, to spell out all of the issues in detail as best we can discern them, so that a case may be made for a particular position in contrast to rival alternatives.

Following this approach, we argued that scholastic aptitude was neither fixed endowment nor direct learning outcome but, rather, developed (and developing) ability. We accepted as a working conception of coaching any intervention procedure specifically undertaken to improve test scores, including training techniques ranging from short-term test familiarization and practice to long-term instruction. Given the diversity of empirical outcomes attendant upon this range of coaching procedures, we felt compelled to ask not whether coaching worked but to what degree, because the averages computed in pursuit of the former categorical question yielded a misleading middle ground uncharacteristic of either short-term or long-term programs. As a consequence, we examined the relationship between the size of score effects attributable to coaching and the amount of student contact time entailed in each coaching program, revealing in the process rank-order correlation coefficients upwards of .7. This relationship proved not to be linear, however, but approximately logarithmic, with arithmetically increasing amounts of score effect being associated with geometrically increasing amounts of student contact time, which in the extant studies may also be serving as a proxy for increasing curriculum emphases on knowledge and skill development. Thus, SAT coaching effects -- within the limitations of the available fragmentary data -- appear to be subject to markedly diminishing returns.

The magnitude of individual score effects, transformed into units of the standard deviation of the score distribution, was interpreted in terms of the amount of improvement a student thereby attained on the average in corresponding percentile rank -- that is, in terms of the percentage of test candidates that a student surpassed by virtue of a particular coaching effect. This improvement in percentile rank, however, is contingent on the student's initial uncoached score, with a given coaching effect yielding its largest percentile improvement for initial scores at the mean of the distribution and falling off as initial scores deviate from the mean in either direction. For example, a 20-point score effect for a student initially scoring one standard deviation below the mean would yield an improvement of about 5 percentiles (from a rank of 16 to 21); if initially scoring one standard deviation above the mean, the improvement would be about 4 percentiles (from 84 to 88); and, if initially at the mean, it would be about 8 percentiles (from 50 to 58).

It is unlikely that percentile improvements of these magnitudes would substantially influence admissions decisions, except possibly for students whose uncoached scores are at the margins of admissibility to selective institutions of their choice. Nor is a coaching effect of this magnitude easily attainable, especially in Verbal. According to the logarithmic model relating student contact time to average score effect based on the available SAT coaching data, about 57 hours of Verbal coaching or about 19 hours of Math coaching would each be expected to be associated with an average of 20 score points or so. Thus, dependent upon the initial uncoached score and its relation to the margins of admissibility to a particular college and noting that geometrically increasing amounts of student contact time are, required for larger and larger score effects, the ultimate value judgment with respect to coaching — Is it worth it? — comes down to the individual case. The evidence and and arguments summarized here offer a rational framework for addressing this issue.

But if obtained percentile increases attributable to coaching represent enduring ability improvement, then the question of worth takes on a different character independent of initial score level or questions of admissibility. The burden would shift to asking when, if ever, ability development was not worth it and how we might incorporate facilitative techniques into the formal educational process.

References

Alderman, D. L., & Powers, D. E. The Effects of special preparation on SAT-verbal scores. *American Educational Research Journal*, 1980, *17*, 239-253. (Also reported as ETS RR 79-1 and CB RDR 78-79, No. 4, Princeton, NJ: Educational Testing Service, 1979.)

Anastasi, A. Abilities and the measurement of achievement. In W. B. Schrader (Ed.), *New directions for testing and measurement -- Measuring achievement: Progress over a decade (Proceedings of the 1979 ETS Invitational Conference)*. San Francisco: Jossey-Bass, 1980.

Bejar, I. I., & Blew, E. O. Grade inflation and the validity of the Scholastic Aptitude Test. *American Educational Research Journal*, 1981, *18*, 143-156.

Carroll, J. B. On the theory-practice interface in the measurement of intellectual abilities. In P. Suppes (Ed.), *Impact of research on education: Some case studies*. Washington, DC: National Academy of Education, 1978.

Churchman, C. W. *The design of inquiring systems: Basic concepts of systems and organization*. New York: Basic Books, 1971.

Churchman, C. W. *The systems approach and its enemies*. New York: Basic Books, 1979.

Coan, R. W. Facts, factors, and artifacts: The quest for psychological meaning. *Psychological Review*, 1964, *71*, 123-140.

Cochran, W. G. The use of covariance in observational studies. *Applied Statistics*, 1968, *17*, 270-275.

Coffman, W. E., & Parry, M. E. Effects of an accelerated reading course on SAT-V scores. Personnel and Guidance Journal, 1967, 46, 292-296.

Cronbach, L. J. Validity on parole: How can we go straight? In W. B. Schrader (Ed.), New directions for testing and measurement -- Measuring achievement: Progress over a decade (Proceedings of the 1979 ETS Invitational Conference), San Francisco: Jossey-Bass, 1980.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods. New York: Wiley, 1977.

Dear, R. E. The effect of a program of intensive coaching on SAT scores (ETS RB 58-5). Princeton, NJ: Educational Testing Service, 1958. (Reported in French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. Educational and Psychological Measurement, 1959, 19, 319-330.)

Dyer, H. S. Does coaching help? College Board Review, 1953, 19, 331-335. (Reported in French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. Educational and Psychological Measurement, 1959, 19, 319-330.)

Educational Testing Service. Test use and validity: A response to charges in the Nader/Nairn report on ETS. Princeton, NJ: The author, 1980.

English, H. B., & English, A. C. A comprehensive dictionary of psychological and psychoanalytical terms. New York: Longmans, Green, 1958.

Evans, F. R., & Pike, L. W. The effects of instruction for three mathematics item formats. Journal of Educational Measurement, 1973, 10, 257-272.

Federal Trade Commission, Bureau of Consumer Protection. Effects of coaching on standardized admission examinations: Revised statistical analyses of data gathered by Boston Regional Office of the Federal Trade Commission. Washington, DC: The author, 1979.

Frankel, E. Effects of growth, practice, and coaching on Scholastic Aptitude Test scores. Personnel and Guidance Journal, 1960, 38, 713-719.

French, J. W. The coachability of the SAT in public schools (ETS RB 55-26). Princeton, NJ: Educational Testing Service, 1955. (Reported in French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. Educational and Psychological Measurement, 1959, 19, 319-330.)

Hammond, K. R. Integrating fact and value in discussions with Ralph Nader. Colloquium presented at Educational Testing Service, Princeton, NJ, April, 1980.

Kaplan, A. The conduct of inquiry. San Francisco: Chandler, 1964.

Marron, J. E. Preparatory school test preparation: Special test preparation, its effect on College Board scores and the relationship of affected scores to subsequent college performance. West Point, NY: Research Division , Office of the Director of Admissions and Registrar, United States Military Academy, 1965.

Messick, S. The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 1975, 30, 955-966.

Messick, S. The effectiveness of coaching for the SAT: Review and reanalysis of research from the fifties to the FTC (ETS RR 80-8). Princeton, NJ: Educational Testing Service, 1980. (a)

Messick, S. Test validity and the ethics of assessment. American Psychologist, 1980, 35, 1012-1027. (b)

Messick, S. Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin, 1981, 89, 575-588. (a)

Messick, S. Evidence and ethics in the evaluation of tests (ETS RR 81-9). Princeton, NJ: Educational Testing Service, 1981. (Educational Researcher, in press.) (b)

Messick, S., & Jungeblut, A. Time and method in coaching for the SAT. Psychological Bulletin, 1981, 89, 191-216.

Pallone, N. J. Effects of short-term and long-term developmental reading courses upon S. A. T. verbal scores. Personnel and Guidance Journal, 1961, 39, 654-657.

Pike, L. W. Short-term instruction, testwiseness, and the Scholastic Aptitude Test: A literature review with research recommendations (ETS RB 78-2 and CB RDR 66-7, No. 1). Princeton, NJ: Educational Testing Service, 1978.

Roberts, S. O., & Oppenheim, D. B. The effect of special instructions upon test performance of high school students in Tennessee (ETS RB 66-36 and CB RDR 77-78, No. 2). Princeton, NJ: Educational Testing Service, 1966.

Slack, W. V., & Porter, D. The Scholastic Aptitude Test: A critical appraisal. Harvard Educational Review, 1980, 50, 154-175.

Snow, R. E. Representative and quasi-representative designs for research on teaching. Review of Educational Research, 1974, 44, 265-291.

Snow, R. E. Aptitude and achievement. In W. B. Schrader (Ed.), New directions for testing and measurement — Measuring achievement: Progress over a decade (Proceedings of the 1979 ETS Invitational Conference). San Francisco: Jossey-Bass, 1980.

Stallings, J. Allocated academic learning time revisited, or beyond time on task. Educational Researcher, 1980, 9(11), 11-16

Stroud, T. W. F. Reanalysis of the Federal Trade Commission study of commercial coaching for the SAT (ETS RB 80-10). Princeton, NJ: Educationa Testing Service, 1980.

Webster's seventh new Collegiate dictionary. Springfield, MA: G. & C. Merriam, 1969.

Whitla, D. K. Effect of tutoring on Scholastic Aptitude Test scores. Personnel and Guidance Journal, 1962, 41, 32-37.

Footnote

[1] An abridged version of this report appears in B. Green (Ed.), Measurement in the Public Arena. San Francisco: Jossey-Bass, 1981. I wish to thank Albert Beaton, Thomas Donlon, Garlie Forehand, Norman Frederiksen, and Robert Linn for their helpful comments on the manuscript and George Miller, Donald Rock, Ledyard Tucker, and Howard Wainer for their advice on various aspects of analysis. Special thanks go to Ann Jungeblut for her pervasive contributions to the substance and phrasing of this work.

Mr. WEISS. Thank you very much. Mr. Erdahl?

Mr. ERDAHL. Thank you very much, Mr. Chairman. Before I direct a question to our witness, I'd like to direct an observation or question to you, Mr. Chairman. I really think you should be encouraged by the response to your proposed legislation. It seems like you have struck the fear of the law givers into the hearts of the test givers. My point would be if the voluntary policies of other testing associations are, I guess I could use the word as conciliatory, as ETS they seem to have negated the real or imagined need for your bill.

Mr. WEISS. If the gentleman will yield.

Mr. ANRIG. I'd be glad to second that motion, Mr. Chairman.

Mr. WEISS. As I indicated at the outset of Mr. Anrig's testimony, I am indeed heartened by the changed attitude. I'm also aware of the fact that had it not been for the potential of legislation being enacted I doubt if we would be at the stage where we are right now. So I think that perhaps legislation and perhaps the enactment of legislation may set them on an even straighter path than they're traveling now.

Mr. ERDAHL. I'm sure that many, as I am, are properly impressed by your legislative clout.

Mr. Anrig, thank you very much for being with us today. I regret that because of a, I might as well drop a name, breakfast meeting with King Hussein, I was not here for the first part of your testimony.

If you haven't done it, could you give us a brief summary of your background for the record because I think it's relative to the position that you've taken today.

Mr. ANRIG. Thank you, Mr. Chairman. Through the Chair to Congressman Erdahl, I've been a teacher, principal, and superintendent of schools at the local level of government. Then for 3 years was in the Department of Health, Education and Welfare, first as the Director of the Division of Equal Educational Opportunities in the Office of Education, and subsequently as Executive Assistant to the U.S. Commissioner of Education. Following that I went to the University of Massachusetts at the State level of government, and from the University of Massachusetts was appointed to commissioner of education for the Commonwealth of Massachusetts in 1973, a position which I've held up to September 1 of this year, when I went to ETS. This is my first experience out of the public sector and out of government. I must admit I'm enjoying it.

Mr. ERDAHL. Thank you. Mr. Chairman, another question. Does ETS have a profile in the New York situation of the students, either academically, income wise, whatever you want to say, a profile of the students that request disclosure? Who are the students that are asking for disclosure anyway?

Mr. ANRIG. Through the Chair, generally with regards to the law school admissions test, you have a very high request for disclosure and I believe it largely cuts across the whole spectrum of youngsters and others who wish to go to law school. In the other areas, however, you've had a much smaller request for participation which either may be a result of the desire for the test or as Congressman Weiss indicated the forms may not be as accessible as the law school forms. But whatever the reason the people who request tend to be from upper income situations. By upper income, apparently it's over the $30,000 a year bracket which is pretty upper.

Mr. **Erdahl**. Mr. Chairman, tomorrow we'll be hearing from some people discussing and I think responding to questions about the impact of testing—not testing, excuse me, coaching. With that in mind, looking at the provisions of H.R. 1662, this requires test agencies to furnish a statement on the extent to which test preparation courses improved test subject scores on the average. Could you comment on that kind of anticipating what we might hear tomorrow?

Mr. **Anrig**. Let me just say, Mr. Chairman, I'll be glad to do that without legislation. We have a document hot off the press, "Issues of Effectiveness and Equity in the Coaching Controversy," it is in my mind the best piece of information I've seen on the issue of coaching. It's comprehensive. It includes all of the previous studies, Porter and Slack aside, it does include all the previous studies. It's been developed by Samuel Messick, the distinguished research scientist from ETS who will be here tomorrow. If the committee wishes to question him, he'll be available to answer your questions. With the Chair's permission, I'll submit it for the record.

Mr. **Erdahl**. I was just going to request that, Mr. Chairman, ask unanimous consent that that be included as part of our record for the committee and staff.

Mr. **Weiss**. Without objection, it's so ordered.

Mr. **Anrig**. Through the Chair to Congressman Erdahl, the issue of coaching to me is debated by researchers and will be debated by them tomorrow, I'm sure. I look at it as an old teacher and I know Congressman you do too. I've never told a youngster not to prepare for a test and never will. You know, it goes against the American dream that practice makes perfect. We all believe that. The important thing isn't whether a youngster should practice for a test or not, of course they should. It's a matter of what they expect to come from that practice. It will not perform miracles and it's very important that they realize that it won't perform miracles. Just as your cramming and mine the night before an examination in college didn't make up for a semester of not doing our work, coaching isn't going to make up for 4 years of not doing your work in high school. The tests measure what you've learned. The aptitudes are learned aptitudes, they're not genetic aptitudes, they're learned aptitudes. So if you practice you'll do better, but if you practice for a short period of time you're not going to do very well. If you practice for a longer period of time, you'll do a little bit better, but it doesn't make that much difference. For instance, if you spend 20 hours of intensive coaching and practice, including the homework that goes with that, you probably can increase your verbal scores about 13 points. If you spend 20 hours on math, a total of 40 hours now, that's a lot of time in addition to going to school. A total of 20 additional study on math you might pull up your score about 21 points. What that really means is that on a scale of 600 points, a lot of people forget that the SAT score is on a scale of 600 points, you might increase 13 points on one hand and 21 points in another, that's two or three questions on the test, that's all. That's worth doing, I'm not going to tell a youngster not to prepare for that, but I want to be sure that he or she understands that it's not going to get him into Harvard. It's the 4-year record that counts in your performance, not the record of what you do in one 3 hour sitting at the test. So I think the issue of coaching from a teaching point of view has become exaggerated beyond its importance. I also think it's important, however, not to deny that there is a role. I think that coaching can take place at home. It doesn't have to take place at school. As I've said in my testimony I'm very determined to provide materials readily accessible to anybody of any economic level so they don't have to feel the only choice is to go to a school and pay $300 worth of tuition.

Mr. ERDAHL. Thank you very much, Mr. Anrig. This may be my final question and it maybe sounds a bit facetious but I don't mean it that way. I would think that ETS could financially and administratively handle the compliances required in this proposed bill, but what about some of the smaller testing agencies? Would we see ETS end up as something of a monopoly in the field.

Mr. ANRIG. Well first, through the Chair, I wouldn't want to see a monopoly out there. I welcome competition We've been winning some of our competition and we've been losing it. We've lost some bids recently. The image of ETS as a monopoly out there is just not true. There's about 3,000 different organizations.

Mr. ERDAHL. I didn't say, sir, it was. I said it could become.

Mr. ANRIG. But you gave me a good opportunity, I wanted to make that point. There are other organizations out there and they're competing with us now. The issue of cost is real. If those organizations can't pass their costs on to the clients then I think that's very true. If they make their costs force the fees to be exorbitant then people will say I'm just not going to take that test anymore. So there is a danger of that. We're not at that point yet.

Let me hasten to add, however, that ETS because it is large should not be viewed as invulnerable to that point. A year ago ETS lost money for the first time in its history. When I was interviewed I said that was a new definition of not for profit. It made me a little nervous. But our margin of revenues over expenditures is a very small percentage. We operate on a fairly small margin. So we can't afford that any more than some of the other organizations can. The main point that you raise which is can these costs in effect affect the welfare of the testing organizations out there, the answer to that is clearly yes.

Mr. ERDAHL. I hope that you're encouraged that your definition position puts you in a unique league with the U.S. Government, the Chrysler Corp., and others. Thank you very much for being with us. Thank you, Mr. Chairman.

Mr. WEISS. Thank you, Mr. Erdahl. Mr. Williams.

Mr. WILLIAMS. Thank you, Mr. Chairman. Mr. Anrig, the New York Truth in Testing Law is entering its third year. How is it working from your viewpoint.

Mr. ANRIG. Congressman, the experience I've had and I want to stress I've only been in office since September so I'm new at this. It seems to me the law is working smoothly because of good faith by all those involved. Good faith by the testing organizations, good faith by the State Commissioner of Education, and good faith by Senator LaValle who is as you may know and I believe a spokesperson is here for him today has amended his law to take care of some of the problems that arose immediately after its passage that everybody recognized wasn't the intent of the law. So there's been a good deal of flexibility both on the law sponsor's side as well as those that are carrying out the law. One of the reasons I think it's important not to legislate at the Federal level is because there is a statute in place which by voluntary decision of the test organizations now applies nationwide, and we ought to just see how that works for a period of time. I think it's working rather well. I think some of the cataclysmic predictions didn't come true. I think some of the miracle predictions didn't come true either. So the best thing to do based on my experience is don't legislate where you're not sure, and it seems to me this is an area we should watch for a while.

Mr. WILLIAMS. Did ETS have a position on the New York proposal prior to its becoming law?

Mr. ANRIG. Through the Chair, the Educational Testing Service was in opposition to the New York State law.

Mr. WILLIAMS. But you now support its implantment throughout the nation on a voluntary basis?

Mr. ANRIG. The decisions of the boards which contract with us, the Law School Board, the Graduate Management Board, and so forth that I mentioned, and the college board, their decisions have been to apply what they are doing in New York State generally to all States so that there's not one State treated differently from the others. There are some exceptions to that but in general that's the position they've taken.

Mr. WILLIAMS. If the various states are allowed to voluntarily abide by the New York State law, won't that present your company and other testing companies with a patchwork quilt throughout the country and make compliance and coordination very difficult?

Mr. ANRIG. With the Chair's position, the decision of the various boards to apply the New York State entitlements nationwide I think largely was based to avoid that. If each State started to pass a New York State-like law then you would find differences because as you well know from your own experience, going though the State legislature someone always wants to put an amendment on that's uniquely theirs. So then you would end up with the patchwork. I think that's not advisable. One way to avoid that would be Congressman Weiss' proposal which is that of a Federal law. I happen to think that the remedy is as bad as the disease. So the way to do it, it seems to me, and it's an obligation on our part to respond is to say we'll respond to the issues that you've raised, that the States are raising and the Federal Government is raising. We'll respond to that and do so voluntarily. It seems to me the record shows that that has been the case.

Mr. WILLIAMS. So the thrust of the New York law and Congressman Weiss proposal is agreed to at least by your testing company, but you'd prefer that that be done—that that thrust be accomplished voluntarily throughout the country rather than by law. Do I understand that as your position?

Mr. ANRIG. Through the Chair I'd like to state the position a little differently than you just did, which is the position of the Educational Testing Service is in favor of openness in testing, we commit ourselves to that, and we take the actions that I've described in my testimony today. To the extent those are complimentary to elements within Congressman Weiss bill, then I'm certainly supportive of those elements which are complimentary but not in general with Congressman Weiss bill, with all respect.

Mr. WILLIAMS. Do we agree that the public has the greatest stake in openness and truth in testing?

Mr. ANRIG. Through the Chair, yes. Affirmatively and because there is a public interest, I believe that organizations like Educational Testing Service have both the same interest and an obligation in that area.

Mr. WILLIAMS. I guess that the question for the committee to decide is given that there is I suppose near unanimous agreement that the public does have this enormous stake in testing in this Nation, the decision then is should we allow private organizations to implant the rules by which truth in testing will be achieved or should that be done by the public through their representatives?

Mr. ANRIG. Through the Chair, Congressman as I said in my opening testimony I've been in government all my life, and I've seen the problem of first trying to legislate a remedy and then administer it afterwards, and I don't have to describe that problem to you, you face it all the time. What you intended isn't necessarily what ends up being. The issue is to look at the results. Are you achieving the results that you want without Government action? If the answer to that is yes, then don't take Government action. If you think you could achieve better results, and that means not just in what you do but in what those who follow you will do also. I say this as a lifelong bureaucrat. Then you look at legislation. But you say boy we might pass a law up here and it'd be a pretty good law and we wanted to do the right thing, but look at those laws we've done the right thing on before and what's happened to them and what ends up being a burden on local people. I just don't think I want to head down that way. So I think that's a very appropriate judgment for the law givers of this country to make. My advice would be that you're achieving your goal without legislation, don't legislate. But that decision rests with the Members of Congress and I recognize and respect that fully.

Mr. WILLIAMS. I appreciate that. I do want to make the point that as a lifelong observer of government I want to make the point that not every problem has been made worse by government involvement. In fact, I really think that a historical overview of the past 50 years shows that most problems government has attacked has been made better by the laws that government has passed. I don't know how political that is to say that in today's climate in which government can do no good. But the hard fact is that government has done an enormous amount of good in the past 50 years and indeed has achieved results that were almost precisely what government intended to achieve.

I, like you, an a former administrator and a teacher, I've been impressed by what can be achieved through testing and equally impressed and concerned with the inexactness of testing. Whether it was grade school or high school, I could move test results simply by moving the day on which I gave the test from Monday to Friday. It seems to me there has been, and I'm going to take one side of this and advocate it and then ask for your response, sir. It seems to me the tremendous focus that we have placed on testing in this Nation has resulted in students or potential students having success or failure based on a test or a very few numbers of tests. For example, admission tests. I'm wondering if you as an expert in this area can speak to whether or not we overuse tests, whether or not we apply them incorrectly, and particularly whether or not we rely heavly on admission test.

Mr. ANRIG. Through the Chair, I think you've raised some very important points, Congressman. Let's take a look at what the result has been. Is it true that some thing out there, maybe tests is stopping students from getting into colleges or getting into the colleges of their choice? I was interested in the letter that was submitted to the committee by the American Council on Education that represents most of the institutes of higher education. According to their data 90 percent of the students currently seeking admission to undergraduate studies are admitted to the college of their first choice. More than 95 percent are admitted to some college. Tests have not stopped people from getting to colleges or even to the colleges of their choice. I think a very strong case could be made quite the opposite, that it has made that possible.

Now, the situations that you describe and both of us know that that does happen, are misuses by individuals of what a test should be. It is true they are inexact, absolutely true; and if anybody were to take a single test as the determining factor for any decision, whether it's promotion or diploma or admission to college or saying the youngster is a good youngster or bad youngster, that's a wrong action. To my knowledge every testing organization that publishes tests takes that position. Now the question is, how can you guard against that kind of abuse. The trouble is you can't do it by law. You can't legislate attitude. An admissions officer who would let a test be the determining factor is not doing his or her job. I say that on the record and anybody else in favor in representing the admission field I'm sure would join me in that position, but there probably are people who do that trying to find an easy way to deal with a large number of people. That's wrong.

Mr. WILLIAMS. Let me interrupt there, sir. Last year I believe it was last year we heard this legislation and a representative of one element of the medical community came before us, I believe in opposition to the chairman's bill, and based his opposition primarily on this point if my memory is correct. Since the use of admission tests the medical colleges with which his association dealt had a significant reduction in the number of drop-outs. On the face of it that sounds good and yet when I asked him if they were relying almost soley on admission tests to reduce that dropout rate, his answer was yes. When I asked if they dealt with the relevancy of curriculum to minorities, his answer was no. And when I asked similar questions his answer was no, we deal just with admission tests. Now that was from a representative of an association that has I suppose thousands of members in the medical community across this country.

Mr. ANRIG. Through the Chair Congressman, there is in the audience and on your schedule of witnesses somebody from the Medical College Admissions Testing program who is better able, I think, to respond to the point you just raised. Let me just answer it more in terms of general questions of admission. Admissions officers in general, and at least I have found especially so in the highly selective institutions, bend over backward to find out additional information about students to that which is provided both by their grades and their scholastic aptitude test scores. I was a little concerned on the recent Phil Donahue Show that you may have seen with Drs. Porter and Slack from Harvard. They were criticizing the SAT and making generalities about admissions. I wish they'd gone right down the street to their own admissions office at that university to see what they were saying is not, in fact, what is done. I have sat in on admission discussions at that institution, and there is an enormous effort to try and find out what underlies the background of a youngster. What more information could we find that makes it clear that this is the youngster we would want have come to this institution. Most admission officers, I think very much, try to do that. Certainly most principals and guidance counselors try to do it. But there is always the aberration. I would hope no institution would take a position that would say that its sole determining action will be based on any test score as a single determining factor overriding any other part of the admissions process.

Mr. Craig. Will the gentleman yield?

Mr. Williams. Just one minute, are those efforts within the admissions office inclusive or exclusive? Are they for the purpose of bringing in the student or excluding those students that may not be fit.

Mr. Anrig. Through the Chair, most institutions now we have to remind ourselves are inclusive institutions. They're not sorting out; they're looking for students. That will increase over the next few years because of the drop in enrollments which is characteristic certainly of your State I believe and others as well. So the number of students available to come to the institutions is going down and will, I believe, about 25 percent nationwide. So institutions are not doing the sorting that may have been characteristic back in the 1950's and 1960's for instance. The small number of institutions that are selective, maybe 50 to 90 institutions in the country, the highly selective institutions are very careful not to rely solely on a test score. They really are very careful not to do that. I just think that we're moving out of the period when that could be a potential danger because there just aren't that many students and schools want to maintain their enrollment.

Mr. Williams. I yield.

Mr. Craig. Yes, Mr. Chairman, I was wondering if it would be possible if we could have the gentleman in the audience, Dr. Cooper, respond at this point specifically to the question of the Congressman.

Mr. Weiss. If you will we have the schedule set up and we'll take it in proper order. Otherwise I think we're really going to dislocate the schedule.

Mr. Craig. OK.

Mr. Weiss. I appreciate the suggestion.

The gentleman from Montana, questions?

Mr. Williams. Mr. Chairman, thank you for your generosity and time.

Mr. Weiss. Mr. Craig?

Mr. **Anrig**. May I, Mr. Chairman, just before we leave the Congressman from Montana, I do want to say that as a lifelong Government servant that I do believe Government has done some good work, particularly in Massachusetts.

Mr. **Williams**. Especially in Massachusetts.

Mr. **Anrig**. I didn't want to in any way deny that. It's just that we've also got some problems. Thank you, Mr. Chairman.

Mr. **Craig**. I'll let you two discuss that. Thank you very much, Mr. Chairman. A couple of questions, I think the thing that concerns me most is that we seem to be saying that your bill and the New York bill are similar and, therefore, certain kinds of things that have resulted from the New York legislation have already come to pass; and, therefore, certain kinds of expenses that would be necessary might not be arrived at with the passing of federal legislation. My quick overview of that finds seven rather major areas of difference, indicating there is a real difference between the New York bill and the proposed legislation we have before us.

One of those areas that concerns me is, and possibly you could address yourself to it, the bill we have before us, in section 4, requires testing agencies to reveal proprietory information in the case of contracts with other agencies, and there's no similar provision in the New York law for that. How does your organization view that particular provision and the potential problems with it.

Mr. **Anrig**. Through the Chair, Congressman the position I would take on it is from a position of an organization which already takes its annual audit report and publishes it. So we put out on the table what our expenditures are, what our revenues are, where they are, and Cooper and Liebring comes in and audits it and we publish it as part of our annual report. We do that by our own decision though. I would find very intrusive Government saying to any private organization, including a not-for-profit organization that you must provide what one could well call legitimate proprietary information and submit that onto the public record. There are organizations, our own included, that have marketing strategies. That if competition were to look at it could pick out those strategies very easily by watching that over a period of time. I'm sure if the Chair were in a law firm and at some future date returned to the halls of Congress and Congress were to say we'd like that law firm of which you were a partner—a high partner I might add—we want you to put your finances out on the public record. The response to that would be pretty strong. I think as a matter of principles that is intrusive. That's Government intrusion into the private sector that I think is not warranged. But again I say that on the basis that as a practice the ETS Board of Trustees has always put on the record its finances. But that is not true in general throughout testing organizations or any other business.

Mr. **Craig**. Mr. Chairman, Mr. Anrig, by my reading of the proposed legislation, are we not saying in that section that if you contracted with a university to do a special kind of study for them, because they were interested in the type of student coming in, should they be different from other universities in certain respects and, therefore, were seeking specialized information, that this would demand your disclosing this information if requested by the public when, in fact, it might be the university's information. Do you see that as a potential conflict?

Mr. ANRIG. Through the Chair, I think the potential is there. What I think the sections, particularly sections 3, 4, 5, 6, 7, and 8 of H.R. 1662, what they do is begin to cut in in a way that I don't believe is true in any other sector of the economy to the really proprietary and private information which an organization, which is private, is entitled to. And I think whether or not it would be true as you just interpreted it, the fact that it raises that question just indicates the problems. How many law cases would be necessary to resolve that matter. How many lawyers would have to be hired to fight that out in court. It's just an area we should get into in my judgment.

Mr. CRAIG. Apparently before I came in, Mr. Chairman, you asked the question as it related to costs and compliance with the New York law and I have read what those answers are. Do you at this time, or has your organization at this time made any evaluation based on potential of additional compliance through the bill that we have before us? What additional costs might be incurred obviously by the consumer?

Mr. ANRIG. Through the Chair, there are some differences between the New York State law and H.R. 1662 and I'm just trying to look at the ones that have a particular cost impact.

Mr. CRAIG. You have done no study at this time?

Mr. ANRIG. Not at this time, Congressman.

Mr. CRAIG. To make a determination as to potential costs?

Mr. ANRIG. We could take a look at that if you'd like and I'd be happy to try to do so.

Mr. CRAIG. I think that would be important for the record.

Mr. ANRIG. Surely. More importantly though the earlier line of questions you had is almost more important than the cost factor, which is the intrusion issue. It would in my judgment treat one sector of the economy, in this area, the private testing development organizations differently than any other sector of the economy is treated by the government. I know of no other sector, other than federal regulated drug companies perhaps, even there I don't believe it's true. I know of no other sector that is treated this way. This is extremely intrusive.

Mr. CRAIG. Thank you very much. I might say, Mr. Chairman, in the overall consideration of this question, at a time when we are making an effort within Government to deregulate, instead of increase regulations, (and Congress has been on record for several years as moving in that direction) that we move out into the private sector when we have large testing activities within Government itself, the civil service type of testing and military testing. There has been real questioning of the related acts that may be unfair to those participants in those tests. Perhaps we ought to be directing ourselves to the problems of Government instead of private industry, and I see nothing in this legislation that encourages that direction or offers any relaxation of potential problems that exist within Government. I see nothing in this legislation that does anything to guard against the institutions' misuse of existing procedure or test information. I think that's another area of concern that, if we're going to go into this, we ought to cover the water-front, as the case may be. Misuse can occur at a variety of levels. Most assuredly it could at the institutional level, dependent upon those making use of the test score itself. Thank you very much for your testimony.

Mr. WEISS. Thank you very much, Mr. Craig.

Just on the last point, without attempting to engage in debate between us, I don't want to tell the institutions how they ought to conduct themselves. I don't want to set educational or admissions policy. I don't think that's appropriate. Indeed what this legislation intends to do, and perhaps in some instances imperfectly, is to provide for disclosure, period. We've tried to limit them to the information surrounding the test zone.

Mr. Anrig, I'm not sure I understood your response to one of the questions Mr. Craig asked you. Let me see if I can phrase the question and your response. Are you suggesting that if the intrusive aspects of this legislation were removed so that the only thing we had left were areas which were clearly regarding information surrounding the test itself and disclosure of the tests and studies relating to the preparation of the test questions and analysis of them—that it would be the scholars and the students period—and we would have the contractual and fee and cost questions and all those other intrusive things aside, are you saying that educational testing service would support this legislation?

Mr. ANRIG. No, Mr. Chairman.

Mr. WEISS. Oh, I thought we had arrived at a point of agreement.

Mr. ANRIG. Good try. No, Mr. Chairman, our position is clear, legislation is not needed. The action is being taken in the absence of legislation, why legislate?

Mr. WEISS. I wanted to set the record straight.

Mr. ANRIG. I appreciate you doing so.

Mr. WEISS. I didn't want you to walk out of here and have your Trustees suggest that you went beyond their position.

On the coaching question that has been alluded to earlier by Mr. Erdahl, you are aware of the fact that before your assumption of the presidency of educational testing the Federal Trade Commission came out with a report in which it concluded that indeed the claim of the coaching schools were more accurate and valid than that of the testing companies as to what the beneficial affects of coaching would be as far as test takers are concerned.

Mr. ANRIG. Mr. Chairman, I'm aware that there was a Federal Trade Commission study done by the Boston regional office which as I understand it was not accepted by the Federal Trade Commission itself. Indeed the Federal Trade Commission or at least the staff of that Commission in Washington raised serious concern about the study that was done in Boston. My understanding is that Miss Lois Pine, a candidate for Lieutenant Governor in Massachusetts, will be here tomorrow and I'm sure she'll take an opportunity to describe what she did and why. The Federal Trade Commission itself has done some further study but my understanding is that nobody has—they're putting a lot of disclaimers, including on the covers of those studies, about their own conclusions. The issue to me is not 13 points, 14 points, or 15 points. That's why I think the coaching controversy is heading in the wrong direction if it goes that way. The issue to me is how can we be sure that a student and his or her parents understand the limits of what short-term coaching can accomplish. Understand the options as to where the information for such coaching, and it doesn't have to take place at a coaching school, it can take place in the youngster's study, or bedroom, or living room. That there is information available to them. And third that we help youngsters who want to prepare for tests to do so and to do so accurately. I'm concerned, for instance, one of the books that recently came out suggested that on the issue of guessing and there is a way to guess on these tests when you narrow it down to two of the four options. It's given bad advice. So what we've done is prepare a brochure which I'll be happy to leave with you today which deals with the issue of guessing. It's all right, it's in the test design.

Mr. **Weiss**. Without objection that, too, is made part of the record.

[The information referred to follows:]

PREPARING FOR TESTS

EDUCATIONAL TESTING SERVICE

There's a mythology about taking standardized admissions tests.

- Some people think the tests measure something innate and that there is no way to raise scores on the exams.
- Others say that all anyone has to do is catch on to the system. With some quick tricks, a test score can be easily raised.

These are the myths. What are the facts?

Educational Testing Service's admissions tests measure the extent to which your reasoning ability and skills with words and mathematical concepts have been developed up to the time you take the test.

These abilities are related to academic success in undergraduate, graduate, and professional schools and develop over a period of years through learning experiences in school, in the family, with friends and associates, and in reading and independent study.

The skills and abilities that are tested tend to grow relatively slowly and at different rates for different people. It is important to remember that these tests are not intended to measure your overall worth. Many other qualities not measured by admissions tests, such as motivation, creativity, and artistic skills, may affect your sense of satisfaction and success in life. All are important.

Will test preparation work?

As educational experience is increased, whether in or out of the classroom, admissions test scores will go up if the verbal and mathematical abilities measured are strengthened. The extent of this development is related, within limits, to the time and effort devoted to it.

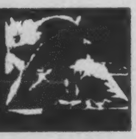
By practicing on ETS sample tests, you will become familiar with the types of questions, the directions, the physical layout of the answer sheet, and you can also practice pacing and gauge your effectiveness at guessing. While such preparation cannot be expected to produce large score gains, being familiar with the test should help you perform to the best of your ability.

Special courses that strengthen verbal and mathematical skills over a period of months may also result in score gains. You should question, however, whether this intense long-term preparation will actually make a difference in your specific plans for college or graduate or professional school and is worth the cost and time to you.

Analyses of all the available research studies on SAT coaching show that approximately 40 hours of class time (spent on both verbal and math studies) plus considerable homework may be expected to yield score gains on the *average* of 13 points in the Verbal and 21 points in the Math sections of the test, depending on the time and effort put in by the student. Each section is scored on a scale of 200-800. These increases correspond to two or three additional correct answers on the 85-question Verbal section and two or three on the 60-question Math section. Within these averages, it is important to note that not all scores rise. Some scores stay the same, some drop, and some are higher.

In preparing for a test, there is no quick remedy for years of neglected studies or weaknesses in the curriculum. The soundest preparation for admissions tests is study in solid academic courses and extensive outside reading.

TIPS FOR TEST TAKERS



ETS wants you to do the best you can on its tests, and offers these tips:

Read the testing information bulletins. ETS bulletins provide a storehouse of information, usually include a full sample test, and tell what to do if you have problems before, during, or after the test. Read the current bulletin of information for the specific test you are taking to learn what is expected, the type of test to be given, its purpose, its format, what knowledge or skills are being tested, and how the scores will be used.

Keep the test scores in perspective. Your test score is used in combination with your high school or college record, recommendations, and other information.

Be prepared. Read all the material about the test carefully, paying particular attention to the directions for each type of question. Take the sample test under timed conditions, and review the questions that you missed. If you want more practice, additional sample materials are available from ETS.

Be alert mentally and physically. Get a good night's sleep before the test date. Arrive at the center in plenty of time. Take extra #2 pencils and identification. Dress comfortably, and eliminate distractions so you can focus on the test. While a little anxiety can add to mental alertness, too much anxiety can have a detrimental effect. Knowing what to expect is the best preparation.

Pace yourself. Bring a watch. Some types of questions are more time consuming than others. Check how many questions there are in each separately timed section and how much time is allowed for answering them. Since most standardized tests give you no more credit for a hard question than for an easy one, skip those that you find are taking too much time. If there is time, you can go back later to questions you have omitted.

Check to see whether guessing is to your advantage. Some tests, like the Scholastic Aptitude Test (SAT) and the Graduate Management Admission Test (GMAT), are scored by subtracting a fraction for each wrong answer from the number of right answers. This process corrects for random guessing. On a test where there is a correction for guessing, guess on those questions where you can eliminate one or more answer choices. For example, if there are four options to a question and you know two of them could not be correct, your chance of selecting the right one from the remaining two is greatly increased.

On other tests, such as the Law School Admission Test (LSAT) and the Graduate Record Examinations (GRE) Aptitude Test, there is no penalty for wrong answers; therefore, try to answer every question.

Fill in your answer sheet carefully. Most multiple-choice tests require you to fill in spaces on a separate answer sheet. Students sometimes mislocate whole blocks of answers on the answer sheet. Check frequently to be sure the number of the answer space you are filling correspnds to the number of the question you are answering. Darken the marks so they completely fill the space; light or partial marks may not be read correctly by the scoring machine. If you change an answer, be sure to completely erase the old choice and mark your new one. Give only one answer to a question. The scoring process for many tests treats two answer marks as an omitted question.

• • •

Useful strategies for test taking include not only general ones, like those above, but some that could help with several specific kinds of questions. Many standardized tests contain antonyms, analogies, sentence-completion questions, and questions based on reading passages. Here are some additional suggestions that could apply to these tests.

TIPS FOR VERBAL QUESTIONS

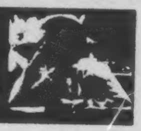
Antonym questions. Read all the choices before deciding which word is the opposite or most nearly the opposite of the word in question. Sometimes it helps to try to provide a context for antonyms. That is, you may not be able to define the word when you come across it standing by itself. So think of a phrase or sentence in which it would be used. If you can provide a context for a word, you may be able to arrive at its meaning.

Analogy questions. Establish a precise relationship between the first two words. Forming a sentence that states the relationship often helps. Be careful to eliminate those relationships that are not exactly parallel to the relationship of the original pair.

Sentence-completion questions. Read the sentence carefully to get a sense of what is being said. In choosing an answer, make sure the word or words are logical within the total sentence. Do not select an answer just because one of the words fits in one phrase or clause. It may help if you fill in the missing words before looking at the answers. Then check the choices to see if any are close.

Reading passages. Read not only for the facts, but for the attitudes presented, logic used, and the basic assumptions presented. Choose your answer on the basis of information in the passage, not your own knowledge or opinion, unless you are instructed otherwise. If a question asks about the main ideas of a passage, do not be distracted by statements that, while true, are only secondary to the main point.

TIPS FOR MATHEMATICAL QUESTIONS



Preparation. If you have not used mathematics for some time, review the basic concepts and vocabulary in arithmetic, algebra, and geometry, or whatever level of mathematics the test covers. The sample tests can be useful for this purpose.

Scratchwork. Space is available in your test booklet for scratchwork. Working a problem out in writing may help you avoid errors in the solution.

Answer choices. In order not to waste time unnecessarily on long computations, check to see if an approximate answer is required; for example, does the question ask for "more than 75 percent" rather than exactly "78 percent"?

Diagrams and figures. If diagrams or figures are not presented, it may help if you draw your own.

Questions based on graphs and tables. Scan the data briefly to see what kind of information is given, and then go immediately to the question(s). You will become familiar with the data gradually as you try to answer the question(s).

Word problems. For mathematical word problems, take one step at a time, reading each sentence carefully and translating the information into questions.

If you are concerned about a test question after the test, contact ETS. We will be glad to consider your ideas and will respond to you.

Refer to your student bulletin for information on how to handle a problem, and if you need more help, contact ETS at the address or telephone number listed in the bulletin or at the ETS general address: Rosedale Road, Princeton, NJ 08541; or telephone: (609) 921-9000.

007453 • S101P1 • 278930 • Printed in U.S.A.

Mr. Anrig. The key issue is the information needs to get out there and get out there more broadly. I've said in my testimony the educational testing service along with its client board will take the responsibility to do that and we are doing that in a lot of different and new formats than have traditionally been available. I'm very anxious, for instance, to get to community agencies on this issue. I think they reach more kids that are poorer youngsters than even the schools do.

Mr. Weiss. You do acknowledge the the information which ETS had provided up to, let's say, 2 years ago in regards to coaching and its effect could not be described as the model of clarity as far as the test taker was concerned as to what he or she could expect from being coached for that exam or SAT?

Mr. Anrig. Again, I want to be clear that some publications that have been criticized are not put out by ETS but rather by those other boards that I've mentioned. But I don't want to walk away from that issue. I think that in general the testing organizations have tended to downplay the effects of coaching. They're not doing that to deny coaching, they're simply saying it's not going to have that much effect. Don't get your hopes up, you can't make up in 40 hours what you didn't do in 4 years. That was their intent in taking that position. In stating that position, they may have overstated it. My own feeling is the debate going on right now is misleading the students. It's giving them hope which is not legitimate hope. The obligation is to make that clear, what it can and can't do. I believe the publication which we have here is extremely clear on that and it traces out the number of hours. If you put in this number of hours in this kind of preparation, you'll get this kind of result for it.

Mr. Weiss. Do you know that it was not until a hearing before this subcommittee about a year ago or so that for the first time ETS explained for the record what it meant as to what kind of coaching would not be helpful, that in fact we're talking short-term coaching and that it defined short term as anything up to 4 hours.

Mr. Anrig. Yes; Mr. Chairman, as I started out my statement, I think the hearings which have been conducted as a result of your bill have been very hopeful in a variety of areas including on the issue of coaching. The problem is to get the information out and to be sure the information is getting out clearly. Through your leadership I think that the information is clearer now than it has been in the past. It's interesting that of the number of press calls I get, I get more on this particular issue than anything else. I think it is getting a lot of national attention and that's good, as long as we get accurate information out and not inaccurate information. I do think it's important to be sure to look at the conditions that were placed both by the Federal Trade Commission itself and subsequently by the regional office on their own study. So I wouldn't take it as gospel. I'd hate to have misinformation getting out now that we're beginning to get clear information out.

Mr. Weiss. Commissioner, thank you very, very much for your testimony and for your patience in responding to our questions. As I said at the outset, I think that the attitude which you evince really is heartening. We look forward to the implementation of those attitudes.

Mr. ANRIG. Thank you, Mr. Chairman, it's been a pleasure to be here. I thank the members of the committee too.

Mr. WEISS. Next is the director of the American College Testing Service, Philip Rever. Will you correct me if I have mispronounced your name? Is it—is it pronounced as Reever or is it River?

Dr. REVER. Mr. Chairman, congratulations, you did a wonderful job.

Mr. WEISS. Yes? OK.

STATEMENT OF DR. PHILIP R. REVER, DIRECTOR, AMERICAN COLLEGE TESTING SERVICE, WASHINGTON, D.C.

Dr. REVER. Mr. Chairman, I am Philip R. Rever and I am the director, not of the entire American college testing program, to my dismay sometimes, but only to its Washington, D.C., office.

I am delighted to appear before the two subcommittees at this hearing because the topic of the hearing would indeed, have a direct bearing on the American college testing program's assessment, a college admission and placement test which is taken by about 1 million students each year. It would affect the tests that we administer or develop for other agencies with which we have contracts for a variety of services. We do administer medical college admission tests though we do not develop their examination, and we have recently signed a contract to develop the law school admission test though we do not administer this examination. However, I will restrict my remarks to the effects of H.R. 1662 on the ACT assessment and in that regard, I have prepared and submitted to the subcommittees and with your permission I would like to have entered into the record a printed statement, but I would prefer to summarize the points that I have made in that statement rather than read it.

Mr. WEISS. That's fine. Without objection, your entire statement will be entered in the record.

[Prepared statement of Philip Rever follows:]

PREPARED STATEMENT OF DR. PHILIP R. REVER, THE AMERICAN COLLEGE TESTING PROGRAM

Mssrs. Chairmen:

I am Dr. Philip R. Rever, Director of the Washington, D.C. Office of The American College Testing Program, Inc. (ACT). Because H.R. 1662, The Educational Testing Act of 1981, would directly affect ACT's Assessment Program, I am pleased to appear before your subcommittees to present ACT's views on the Bill[1]. In short, ACT opposes passage of H.R. 1662 just as it opposed passage of H.R. 1662's predecessor, H.R. 4949, during the 96th Congress.

ACT's reasons for opposing H.R. 4949 were fully discussed in a comprehensive statement submitted to the Subcommittee on Elementary, Secondary, and Vocational Education on September 10, 1979. Nothing has occurred since ACT submitted its earlier statement that has changed ACT's views about the value, necessity, or educational consequences of bills like H.R. 4949 on the ACT Assessment and students. To the contrary, ACT's experiences in complying with the New York law have reinforced its concerns about the utility and desirability of legislation modeled after the New York law, e.g. H.R. 1662. In addition, the actions taken by test agencies to respond to proponents' concerns about tests and test use in admissions have made H.R. 1662 unnecessary in the absence of substantial numbers of conflicting state laws that regulate test agencies. To demonstrate that passage of H.R. 1662 is ill advised and unnecessary, the experiences of ACT in New York, the changes in ACT policies governing public access to the Assessment, and the changes in hand scoring policies made by ACT since the effective date of the New York law will be reported and discussed in this statement.

---

[1] The Assessment Program is the formal name of ACT's college admission and placement services. The Assessment which is taken by about one million college bound students is the formal name of the test that is the foundation of ACT's services.

ACT's Experience in New York.

In testimony before the Subcommittee on Elementary, Secondary, and Vocational Education on September 10, 1979, ACT reported that compliance with the New York law would require a reduction in test dates, or test availability, in New York, thereby seriously inconveniencing students. In addition, ACT predicted that advantaged students were more likely than disadvantaged students to obtain released test materials, a result of the legislation that was clearly unintended. Evidence gathered to date suggests that ACT's concerns were well founded with regard to the Assessment. ACT's experiences are summarized below:

1. Because of the limited availability of tests for release to the public as required by the amended New York law, ACT reduced its test dates in New York from 9 to 6.
2. The reduction in test dates has been accompanied by a reduction in the number of students who take the ACT Assessment in New York. Since the effective date of the New York law, approximately 4,000 fewer students have taken the Assessment each year than during the test year preceeding the effective date of the law. These 4,000 students did not receive useful and unique data about themselves.

For example, they did not receive an interest profile which is a integral part of the ACT Assessment and which is useful in planning for college and careers.

3. ACT will spend several hundreds of thousands of dollars to increase test production to replace tests retired at an accelerated rate as required by the New York law.
4. Since the effective date of the New York law, very few students have requested and obtained copies of their Assessment test materials. Specifically, less than .2 percent of 1 percent (115 out of 79,723 students who have taken the Assessment in New York since January 1, 1980) have requested and received copies of their Assessment test materials. To paraphrase Winston Churchill, never has ACT expended so much for the benefit of so few. More importantly, the extremely low demand for Assessment materials caused ACT to carefully consider the appropriate use of student fees to support changes in existing policies and procedures.
5. The 115 students in New York who have ordered copies of their Assessment test materials are not representative of ACT tested students. For example, their average ACT composite test score was 24.2 on a scale of 1 to 36 with an average score of about 19, 85.9 percent reported having a B or higher grade average in high school, 54.3 percent reported

their family's income as $20,000 per year or higher.[2] New York students who requested copies of their test materials were decidedly advantaged according to traditional indicies of relative "advantage."

After repeated assurances from witnesses during previous legislative hearings that student demand for access to test materials was very strong, the exceedingly small proportion of New York students who have requested copies of their Assessment materials may be surprising to some. ACT was not surprised.

ACT expected little demand in New York and would expect little nationwide demand for access to their Assessment materials were H.R. 1662 enacted because relatively few students, about 4 percent of all tested students, take the Assessment more than once, i.e. "retake" the test. Given the relatively few students who retake the Assessment, the low demand for disclosure, i.e. access to test materials shortly after tests are administered, by New York students was expected.

Low but higher demand for disclosure by New York students was reported by the College Board in previous testimony. Because students believe, often mistakenly, that their performance on the SAT determines their admissability to colleges, substantially greater proportions of students take the SAT more than once (either by taking the PSAT before they take the SAT or by retaking the SAT). Based on the greater proportion of students who retake the SAT, it would be reasonable to expect greater proportions of New York students who have taken the SAT to obtain copies of their SAT test materials. To avoid misunderstanding, it should be made clear that according to figures reported during the previous hearings on H.R. 1662, the proportion of New York students who obtained copies of their SAT test materials was small, about 5 percent, but substantially higher than comparable figures reported by ACT although the procedures students follow for obtaining copies of test materials were similar. Clearly, differences in student demand to review copies of test materials are directly and primarily related to students' perceptions about the use of particular tests in admissions decisions.

---

[2] Comparable data for all New York tested students are as follows: their average ACT composite score is 20.0, 56.0 percent report a B or higher average high school grade, and 36 percent a family income of $20,000 or higher.

In summary, ACT's experiences in New York have tended to substantiate its concerns about the New York law and legislation such as H.R. 1662. To the detriment of tests, test agencies, students, and colleges; such legislation does not respect the substantial differences among tests, test agencies, test usage, and student demand for disclosure. Therefore, passage of H.R. 1662 is ill advised.

Changes in ACT Policies Since January 1, 1979.

While ACT's experiences in New York support its view that enactment of H.R. 1662 is ill advised, ACT has taken actions that make H.R. 1662 unnecessary. Like other test agencies, ACT has (a) modified its policies governing public access to the Assessment and (b) expanded its scoring services as reported below.

ACT has modified its policies governing public access to the Assessment, just as other test agencies have, according to its test development capabilities, the psychometric foundations of its test, the primary use of its test, and student demand for the disclosure as demonstrated by New York students. Prior to September 1, 1981, ACT routinely provided copies of the Assessment to scholars and college faculty members upon their request. Beginning this Fall, ACT also will:

1. send, at the beginning of each test year, a copy of a recently retired ACT Assessment to each high school that received one or more test score reports during the previous test year and to any other high school that requests a copy of the test.
2. invite students to order up to two recently retired forms of the ACT Assessment.

Both changes reflect ACT's experiences in New York, ACT's retest rate, and the concerns ACT has raised about H.R. 1662 and similar legislation in previous legislative hearings. Under these changes students will have access to retired forms of the Assessment before they take the test. If there are benefits to be derived from receiving retired copies of the Assessment, these benefits will be realized before the tests are first taken. ACT will not encourage or cause an increase in unjustified or unnecessary retesting, a possible consequence of the adoption of more extensive disclosure policies.

Because the two forms made available to students will be different forms than the one sent to high schools, and because all three forms will be ones recently retired, all forms of the Assessment will be available for public review after they are retired. Consequently, the public will eventually have access to all forms the the Assessment.

In addition to providing greater public access to the Assessment, ACT has expanded its hand scoring services. Prior to September 1, 1981, ACT hand scored upon request, and at no charge to students, answer sheets in its Iowa City, Iowa headquarters. Beginning in the Fall of 1981, ACT will, upon request, provide hand scoring services at locations where students can conveniently observe and verify their scores. Should the hand scoring be done in ACT's Iowa City headquarters or in one of ACT's 13 Regional offices, no fee will be charged to students. Hand scoring at other locations will be arranged at the request of students although an appropriate fee may have to be charged if additional costs are incurred. Accordingly, beginning this Fall, hand scoring services have been expanded without incurring substantially increased costs that are disproportionate to the demand for score verification as experienced in New York.

Although the changes reported above may be characterized as expansions of existing policies, they were not made hastily. They reflect ACT's concerns about the consequences of disclosure on tests, test usage, students, and postsecondary institutions. Similarly, actions other test agencies have taken in response to the New York law and previous state and federal legislative initiatives reflect the psychometric characteristics of their tests, the uses of their tests, and the demand of students for their test materials. Consequently, passage of H.R. 1662 is unnecessary.

The Future.

Just as ACT has modified and expanded its policies governing public access to the Assessment and its hand scoring services according to its experiences in New York and its best judgment about ways of responding to proponents' concerns about tests and their uses, other changes such as accelerating public access to retired forms of the Assessment will remain under study. Such changes will be made if ACT is confident they will achieve their intended purposes without (a) jeopardizing the quality, value, and availability of the Assessment, (b) burdening students with additional costs for unwanted services, or (c) adversely affecting educational or admission processes. Using the same criteria, ACT has a lustrious history of responding at its own initiative to concerns about testing and improving tests and their uses. The recent changes in policies as reported in this statement are the most recent additions to this history and provide the foundation for future changes in policies and programs.

Because H.R. 1662 is ill advised and unnecessary, ACT respectfully urges that the Bill not be passed. I will be pleased to answer any questions the members of your subcommittees may have about ACT and its position on H.R. 1662.

Dr. **Rever.** In short, the American college testing program opposes the passage of H.R. 1662 for the same reasons it has opposed similar legislation introduced in 23 other States and in the 96th Congress.

ACT views H.R. 1662 as both ill advised and unnecessary. It is ill advised because it mandates uniform test disclosure policies on all test agencies without regard to the substantial differences among tests in their psychometric foundations, their purposes and uses, and the student demand for disclosure as demonstrated in New York.

H.R. 1662 has been made unnecessary in our view by the voluntary efforts of test agencies to respond to the concerns about tests, about test uses, and test agencies in previous legislative hearings.

Let me briefly summarize ACT's experiences in New York that have reinforced ACT's opposition to H.R. 1662.

First, you should know that ACT has complied with the New York law within its capability to do so which is severely constrained by the number of available tests to be made—to be disclosed at an accelerated rate as required by the New York law. Since the New York law became effective on January 1, 1980, 79,723 students have taken the assessment. Of the 79,723 students less than two-tenths of 1 percent, or to be exact, 115 students have requested and received copies of their assessment results and materials. The students who have requested copies of their assessment materials in New York are clearly advantaged according to traditional indices of relative advantage, and I have provided those incidents in my prepared statement.

After hearing repeated assurances from witnesses in previous legislative hearings that student demand for disclosure would be very strong, some may be surprised at the low demand for disclosure for ACT assessment materials. To the contrary, ACT was not surprised because only 4 percent of the students who take the assessment each year take it on more than one occasion, that is they retake the test. This low retest rate can be compared to higher but low retest rates of higher retest rates of other testing programs. The low request rate for ACT assessment materials in the State of New York can be compared to higher but low request rates reported by the college board during the July hearings on H.R. 1662.

Since both the college board and ACT use similar procedures for informing students about disclosure and for students to follow in obtaining copies of their test materials and because other test agencies have reported quite different request rates, notably the law school admission council, it is clear that the dominant factor that determines request rates for test materials are the student perceptions about the use of particular tests and admissions decisions.

In summary, ACT's experiences in New York confirms its opinion that H.R. 1662 is ill-advised.

To move to another point, ACT believes that H.R. 1662 has been made unnecessary by the voluntary actions of test agencies to respond to the concerns that have been raised by the bill's proponents during previous hearings. In fact, in light of its experiences in New York, ACT has modified its policies governing public access to the assessment, and we have expanded our hand-scoring services that have been made available to students.

Prior to September 1, 1981, ACT routinely made available copies of the ACT assessment to scholars and college faculty members for their use and study. Beginning in the fall of 1981, this fall, ACT also will send one copy of a recent retired assessment to each high school in the country that received a score report the preceding test year and to other high schools that request a copy of the instrument. ACT will also invite students to order up to two copies of recently retired forms of the assessment.

These changes mean that all forms of the ACT assessment will eventually be made available to the public. Beginning in the fall of 1981, ACT will expand its hand-scoring services to provide scoring in locations students find convenient for observing and verifying the scoring of their answer sheet. In previous years ACT has hand-scored at no charge and upon the student's requests answer sheets in its Iowa City, Iowa headquarters. Beginning this fall, ACT will offer the same service at its Iowa City headquarters and will offer that service through its 13 regional offices, and scoring will also be conducted at other locations as requested by the students although a small fee and an appropriate fee may be charged should costs be incurred in making those arrangements.

In summary, we believe that the bill before these two subcommittees is ill-advised and unnecessary, and let me just speak for a moment about the future.

Just as ACT has modified and expanded its policies governing public access to the assessment and its hand-scoring services according to its experiences in New York, and its best judgment about ways of responding to proponent's concerns about tests and their uses, other changes such as accelerating public access to retired forms of the assessment will remain under study. Such changes will be made if ACT is confident that they will achieve their intended purposes without (a) jeopardizing the quality, value, and availability of the assessment, (b) burdening students with additional costs for unwanted services, or (c) adversely affecting educational admissions processes.

Using the same criteria, ACT has an illustrious history of responding at its own initiative to concerns about testing and improving tests and their uses. The recent changes in policies as reported in my prepared statement are the most recent additions to this history and provide a foundation for changes in the future.

I would be pleased to respond to the questions that you and your members may have, Mr. Chairman.

Mr. WEISS. Dr. Rever, thank you very much. Again, for the record, let me see if I can get from you certain factual information.

What is the number of students annually who take the—the admissions test—the post-secondary admission test?

Dr. REVER. This past test year, I think it was just a few students over 1 million students took the ACT assessment nationwide.

Mr. WEISS. And would you have any information as to how much of an overlap there is? What percentage do you think, or do you know, of your students also are required to—or do, in fact, take the standard aptitude test?

Dr. REVER. Mr. Chairman, that information is quite elusive despite our best efforts to obtain it, and we've not been successful in doing that. There obviously would be some overlap, but primarily those students who would take both examinations are those who probably are located in different parts of the country applying to a number of different institutions that would require one test over the other or vice versa. For example, as you may recall from our previous testimony, ACT tests roughly 35,000 students in the State of New York. I would guess that many of those students may indeed also take the scholastic aptitude test, though I don't know what that proportion may be simply because they may be applying to institutions in the State of New York which would typically require the scholastics aptitude test, but at the same time they may be submitting an application to a school say in New Mexico in which the typical institution will require the ACT.

Mr. **Weiss**. Again, for the record, in very brief shorthand terms, can you give us a description of the differences between your test, the American college assessment test, and the scholastic aptitude test?

Dr. **Rever**. I can be very brief. We believe ours is better.

Mr. **Weiss**. That's not a description.

Dr. **Rever**. I'm sorry. Our examination was developed initially in 1959 and it has four subparts. It is characterized by the Burrough's Mental Measurements Yearbook as an achievement examination that is reflected in the title of the its four subtests. We have one subtest on English usage, one subtest on mathematics usage, one on natural sciences, and one on social studies. The examination was not intended to be based on experiences of the college board and ETS in their examination was not intended to be a so-called aptitude examination, and I think the kinds of examination item that appear on the examination would be judged by most observers as highly related to the kinds of curriculum content in high schools and are less abstract, if you will, than the scholastic aptitude items.

Now, you know, that those kinds of statements are overgeneralizations obviously, but I think there are some recognizable differences between the examinations if they're laid side by side and one can review them, and that's quite possible.

Mr. **Weiss**. Have you done any studies or has anybody done any studies for correlation as to the scores on those tests and performance records in college?

Dr. **Rever**. Indeed we have. We have probably some studies, and I will be happy to make them available to the subcommittees on the relative predictive validity of our examinations. In other words, we have some institutions which, and this may in fact for your information, be a pattern that has been emerging for a number of years, in which institutions will accept results from either the ACT assessment or the college board scholastic aptitude test for admissions purposes. In some cases we have identified institutions in which there were sufficient quantities of data available for students who submitted scores for the SAT and for the ACT Assessment in which regression equations or predictive validity could be established. In essence, I can summarize those for you.

Both examinations studies the predictive validities of both examinations produce comparable results. In other words, both examinations are about equally predictive of first year college grades. That may vary from institution to institution, and I've given you a generalization, but I can provide you the data we have as a result of our studies.

Mr. **Weiss**. Whatever materials you can provide would be welcome. Have you submitted to us a retired test for purposes of inclusion?

Dr. **Rever**. I neglected to bring some but I will see that they are delivered to the subcommittees tomorrow.

Mr. **Weiss**. They will be entered into the record.

[The information referred to follows:]

Sample Test Booklet*

The American College Testing Program

The ACT Assessment

©1981 by The American College Testing Program. All rights reserved
*Answer key for this sample test booklet is provided on page 35.

8120C

TEST 1
ENGLISH USAGE

40 Minutes—75 Questions

DIRECTIONS: In the passages that follow, certain words and phrases are underlined and numbered. In the right-hand column you will find alternatives for each underlined part. You are to choose the one that best expresses the idea, makes the statement appropriate for standard written English, or is worded most consistently with the style and tone of the passage as a whole. Choose the alternative you consider best and blacken the corresponding space on your answer sheet. If you think the original version is best, choose "NO CHANGE." Read each passage through once before you begin to answer the questions which accompany it. You cannot determine most answers without reading several sentences beyond the phrase in question. Be sure that you have read far enough ahead each time you choose an alternative.

Passage I

Many people think of the desert as barren and desolate, but a visit to Death Valley dispenses[1] such an opinion. Death Valley National Monument offers as many attractions for the tourist than any scenic area.

The valley center consists of a salt lake that had[3] almost dried up, close inspection of the lake reveals salt crystals in infinite variations. Driving to Dante's Point from the lake, a half hour's drive, where[4] you can view the highest and lowest points in the continental United States. In an hour's drive, you can visit a ghost town, hike up Telescope Peak, exploring[5] any of the colorful canyons, or climb down into a volcanic crater. A short trip will take you to the castle constructed by millionaire Albert Johnson. Built in the middle of the desert, the castle took ten years to complete. Although[6] Death Valley remains open all year, the summer months extending from May through October, is[7] too hot to be enjoyable. Despite this extreme heat and lack of rain,

1. A. NO CHANGE
   B. disposes
   C. dispels
   D. impels

2. F. NO CHANGE
   G. from
   H. as
   J. over

3. A. NO CHANGE
   B. has
   C. had been
   D. was

4. F. NO CHANGE
   G. When you drive to Dante's Point from the lake, a half hour's drive,
   H. A half hour's drive from the lake takes you to Dante's Point, where
   J. After you take a half hour's drive to Dante's Point from the lake,

5. A. NO CHANGE
   B. explore
   C. you cam explore
   D. can explore

6. F. NO CHANGE
   G. (Do NOT begin new paragraph) Because
   H. (Do NOT begin new paragraph) Even though
   J. (Begin new paragraph) Although

7. A. NO CHANGE
   B. months, extending from May through October, are
   C. months, extending from May through October are
   D. months extending from May through October is

various forms of life flourish; the many plants and animals living in Death Valley demonstrates[8] Darwin's theory that those who survive must adapt to his[9] environment.

8. **F.** NO CHANGE
   G. demonstrated
   H. will demonstrate
   J. demonstrate

9. **A.** NO CHANGE
   B. its
   C. Darwin's
   D. their

Passage II

Every society has customs and beliefs that are meaningful to some people,[10] irrational to others. These notions, frequently called superstitions, are often based on fear of the unknown or on religious beliefs. Nevertheless,[11] it is considered bad luck to spill salt inasmuch as[12] Judas spilled it at the Last Supper. Crossing one's fingers are physical acts[13] acknowledging the existence of a supernatural being.

The derivation of[14] most superstitions has been forgotten. People who engage in superstitious practices today do so without knowing quite why. Most people do not seriously believe that breaking a mirror will bring seven years[15] bad luck, but they accept this superstition in a spirit of fun. This superstition, however, originated in an age when a reflection is[16] considered part of the soul. Mirror breakage was thus

10. **F.** NO CHANGE
    G. people. And
    H. people,
    J. people; or

11. **A.** NO CHANGE
    B. In addition,
    C. For example,
    D. OMIT and begin sentence with *It*

12. **F.** NO CHANGE
    G. due to the fact that
    H. because
    J. for the simple reason that

13. **A.** NO CHANGE
    B. is a physical act
    C. represents physical acts
    D. represent a physical act

14. **F.** NO CHANGE
    G. (Begin new paragraph) Deriving
    H. (Do NOT begin new paragraph) The deriving
    J. (Do NOT begin new paragraph) Deriving

15. **A.** NO CHANGE
    B. years' of
    C. years of
    D. year's of

16. **F.** NO CHANGE
    G. being
    H. will be
    J. was

GO ON TO THE NEXT PAGE.

1

a serious accident. Even the ritual of wearing a wedding ring has a basis in superstition.

A woman wears a ring on the third finger of her left hand, it's[17] an ancient belief that a vein runs from a[18] finger to the heart. The wedding ring still symbolizes the giving of the heart with the hand.

17. A. NO CHANGE
B. hand. Because of
C. hand because of
D. hand, which is

18. F. NO CHANGE
G. that
H. one
J. your

Passage III

These paragraphs may or may not be in the most logical order. The last item will ask you to choose the most logical order.

(1)

The formal tradition started in the eighteenth century, when slave poets began to use contemporary forms to write about realistic subjects. It continued into the nineteenth century, as black poets added their voices to the clamor for abolition. At the turn of the century, Paul Laurence Dunbar was writing lyrics, for capturing[19] aspects of folk life and that were often in dialect.

19. A. NO CHANGE
B. lyrics, to capture
C. lyrics that captured
D. lyrics, that captured

(2)

Two different traditions have contributed to the development of modern black poetry. The first of these, the folk tradition begins[20] in the seventeenth century, shortly after Africans were first brought to

20. F. NO CHANGE
G. tradition, beginning
H. tradition has begun
J. tradition, began

GO ON TO THE NEXT PAGE.

America similar to slaves. Combining their memories of Africa with their experiences in America, slave singers composed various songs and spirituals to express their feelings about slavery.

(3)

The blending of these traditions can be seen in much twentieth-century black verse. James Weldon Johnson praised his anonymous predecessors in "O Black and Unknown Bards." Containing many allusions to the spirituals, this poem is a fine example of the use of folk materials by a formal poet. During the Harlem Renaissance of the twenties, black poets used techniques and subjects from folk life essentially for their poetry.

Thus, jazz and blues rhythms, ghetto experiences, and black speech has become vital elements of this poetry.

21. **A.** NO CHANGE
B. as
C. in the manner of
D. just like

22. **F.** NO CHANGE
G. Blending these
H. The traditions
J. Such traditional blending

23. **A.** NO CHANGE
B. poet, during
C. poet during
D. poet of

24. **F.** NO CHANGE
G. as essential elements of
H. for essential elements to
J. for the elements which are essential to

25. **A.** NO CHANGE
B. However,
C. Nevertheless,
D. OMIT and begin sentence with *Jazz*.

26. **F.** NO CHANGE
G. will become
H. became
J. have became

27. Choose the sequence of paragraph numbers that will make the essay's structure most logical.
A. NO CHANGE
B. 1, 3, 2
C. 2, 1, 3
D. 2, 3, 1

Passage IV

Jane Austen's reputation has grown so steadily, that the bicentennial celebration of her birth confirmed her

28. **F.** NO CHANGE
G. steadily
H. steadil
J. stead

GO ON TO THE NEXT PAGE.

1

fame as a writer not just for an age, but for all time. Because her six novels were published anonymous, [29] her obscurity during her lifetime was almost complete and, [30] two centuries later, she is cherished as few writers have ever been.

Born in the year of the American Revolution, she too was revolutionary in being England's first lady of literature. Due to succeeding, [31] she charted a new course for the novel of realism, her contributions outdistancing even Sir Walter Scott. [32] By the twentieth century, she became the object of a cult. Kipling's term for such cultists were [33] Janeites.

Janeites consider Austen no less a writer rather [34] an issue. Those readers who [35] fail to appreciate her are looked upon as deficient. Not to like Jane is to be personally inadequate and to lack breeding. Therefore, a reader is compelled to make not solely a literary judgment, but instead [36] a character revelation.

29. A. NO CHANGE
    B. anonymous (without her name),
    C. unknown,
    D. anonymously,

30. F. NO CHANGE
    G. complete, now
    H. complete. Although
    J. complete. Now,

31. A. NO CHANGE
    B. Due to the fact of her success,
    C. Because of the fact of her success,
    D. Because of her success,

32. F. NO CHANGE
    G. those of Sir Walter Scott.
    H. that of Sir Walter Scott.
    J. these of Sir Walter Scott's.

33. A. NO CHANGE
    B. was
    C. are
    D. was to be

34. F. NO CHANGE
    G. then
    H. than
    J. but

35. A. NO CHANGE
    B. readers, who
    C. readers, which
    D. readers whoever

36. F. NO CHANGE
    G. but also
    H. but in addition
    J. as well as

GO ON TO THE NEXT PAGE.

Passage V

Comic books may be the most widely read mass medium in the United States. From their beginning in the 1930s (prior to World War II), 37 comic books have appealed both to young readers and to adults. Today millions of comic books are sold each month, with a "pass on" probable 38 readership of at least three for each individual 39 copy sold.

Comic books in their familiar, magazine, 40 format did not appear until 1933, when 41 *Funnies on Parade* was conceived. This magazine consisted of original 42 reruns from Sunday newspaper comic strips and sold 43 for ten cents. They're 44 instant popularity prompted the publication two years later of comic books to use 45 original material.

After 1935, comic book publication began to undertake 46 tremendous proportions. In 1938, a personality named Superman appeared in *Action Comics*.

37. A. NO CHANGE
B. 1930s (four decades ago),
C. 1930s (when they originated),
D. 1930s,

38. F. NO CHANGE
G. probable (Place after *with a*)
H. probably (Place after *with a*)
J. probable (Place after *readership*)

39. A. NO CHANGE
B. and every
C. purchased
D. OMIT

40. F. NO CHANGE
G. familiar, magazine
H. familiar magazine,
J. familiar magazine

41. A. NO CHANGE
B. 1933, which was when
C. 1933. When
D. 1933, in which

42. F. NO CHANGE
G. all new
H. only original
J. OMIT

43. A. NO CHANGE
B. strips selling
C. strips which sold
D. strips, they sold

44. F. NO CHANGE
G. The comic strips'
H. Its
J. A magazine's

45. A. NO CHANGE
B. using
C. to utilize
D. and using

46. F. NO CHANGE
G. to assume
H. to take on
J. OMIT

GO ON TO THE NEXT PAGE.

1

Superman set the stage for a large range of heroes, with superhuman attributes.

When publishers began specializing in horror and crime comics, many parents became disturbed. Consequently, the comic book publishers organized a comics code authority for review and judging all material intended for publication. However, those publishers who would not abide by the code continued to print comic books of questionable taste.

The value of comic books as an effective medium of communication, is great. Even the United States government uses the comic book format in our drive against narcotic addiction, and many private agencies publish comic books to teach scientific, historic, and moral principles.

47. A. NO CHANGE
    B. heroes, having
    C. heroes which have
    D. heroes with

48. F. NO CHANGE
    G. for review and to judge
    H. to review, and to judge
    J. to review and judge

49. A. NO CHANGE
    B. communication are
    C. communication, are
    D. communication is

50. F. NO CHANGE
    G. her
    H. one's
    J. its

51. A. NO CHANGE
    B. for the purpose of teaching
    C. in order to teach people
    D. so they can teach

Passage VI

In the December cold, a man walks along the beach and studies the sand in front of him. Winter is an unusual time to be strolling the Atlantic beach of New Jersey, the reason the man is there is even more unusual.

Though he carries no metal detector or underwater equipment, this man is a

52. F. NO CHANGE
    G. Jersey and
    H. Jersey, but
    J. Jersey, however

53. A. NO CHANGE
    B. However,
    C. Nevertheless,
    D. Even although

GO ON TO THE NEXT PAGE.

treasure hunter. By the[54] stretch of New Jersey shoreline, Spanish pieces of eight has washed[55] up after winter storms. Nonetheless,[56] there is a treasure trove somewhere just offshore, and[57] the wind and wave action of gales brings it[58] ashore bit by bit.

Because of its cyclonic nature, the wind here generally backs around to the west after a gale from the northeast. This brisk west wind blows[59] sand away from the sides of a coin, building it up into a tiny column and making[60] the coin stand out in relief. To the careful observer, the telltale circle in the sand can mean treasure or a bottle cap or rubber washer.

Treasure hunting[61] is not unrewarding along this coast. Ninety years ago two men rowed ashore from a sailing vessel anchored to[62] a New Jersey inlet and later returned to their ship with an ironbound chest, leaving a scattering of Spanish coins around a hastily dug hole. Since then, anyone with the eyes to see them, has[63] been

54. F. NO CHANGE
G. Up that
H. On this
J. Around a

55. A. NO CHANGE
B. have been washing
C. had washed
D. were washing

56. F. NO CHANGE
G. Nevertheless,
H. However,
J. Apparently,

57. A. NO CHANGE
B. offshore,
C. offshore, and due to
D. offshore and by virtue of

58. F. NO CHANGE
G. them
H. these
J. some of them

59. A. NO CHANGE
B. has blown
C. blew
D. had blown

60. F. NO CHANGE
G. make
H. is making
J. made

61. A. NO CHANGE
B. (Begin new paragraph) The hunting for treasures
C. (Do NOT begin new paragraph) Finally, treasure hunting
D. (Do NOT begin new paragraph) The hunting for treasures

62. F. NO CHANGE
G. from
H. off
J. around

63. A. NO CHANGE
B. them has
C. them, have
D. them have

GO ON TO THE NEXT PAGE.

picking up silver coins right at there 64 feet. However, no one has managed to garner more then 65 a few coins at a time. Nevertheless, the treasure walkers of the New Jersey beaches will undoubtable 66 keep hunting.

64. F. NO CHANGE
    G. his
    H. they're
    J. anyone's

65. A. NO CHANGE
    B. little more than
    C. more than
    D. no more then

66. F. NO CHANGE
    G. with doubt
    H. in doubt
    J. doubtless

Passage VII

Becoming 67 a student at a metropolitan university 68 is to enter a city within a city, a kind of scale-model urban sprawl with its own teeming, heterogeneous population—and its own scale-model urban problems. Parking for instance, 69 is nearly impossible. Every morning, cars cover both legal and illegal parking areas. Students crowd the walks, each person scurrying in various directions. 70

Therefore 71 plagued by the typically urban problems of overcrowding and confusion, the campus differs from the city at large in at least one important way, on campus, there is 72 still traces of

67. A. NO CHANGE
    B. To become
    C. When becoming
    D. When one becomes

68. F. NO CHANGE
    G. student, at a metropolitan university,
    H. student at a metropolitan university,
    J. student, at a metropolitan university

69. A. NO CHANGE
    B. (Do NOT begin new paragraph) Parking, for instance,
    C. (Begin new paragraph) Parking for instance,
    D. (Begin new paragraph) Parking, for instance

70. F. NO CHANGE
    G. a separate direction.
    H. separate directions.
    J. every direction.

71. A. NO CHANGE
    B. Besides
    C. Though
    D. Since

72. F. NO CHANGE
    G. way; on campus. there is
    H. way: on campus, there are
    J. way on campus; there are

GO ON TO THE NEXT PAGE.

an old-fashioned sense of community. Here strangers seem to appear as though they were [73] less suspicious of one another than [74] busy city streets. Newcomers need not be timid about asking for directions, though hurried, people [75] are happy to help, even though no one knows for sure where anything is beyond his own "quad" or street.

73. A. NO CHANGE
    B. to appear as if they were
    C. to give the impression of being
    D. OMIT

74. F. NO CHANGE
    G. then
    H. then on
    J. than on

75. A. NO CHANGE
    B. directions (though hurried) people
    C. directions, though hurried people,
    D. directions. Though hurried, people

END OF TEST 1

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 2

MATHEMATICS USAGE

50 Minutes—40 Questions

DIRECTIONS: Solve each problem, choose the correct answer, and then blacken the corresponding space on your answer sheet.

Do not linger over problems that take too much time. Solve as many as you can; then return to the others in the time you have left.

Note: Illustrative figures are not necessarily drawn to scale and are assumed to lie in a plane unless otherwise stated.

DO YOUR FIGURING HERE.

1. To buy a new car, Joan borrowed $500 at a yearly interest rate of 8 percent. Since there was no early-payment penalty, she repaid the loan in one lump sum at the end of 6 months. How much simple interest did she pay?

A. $ 20
B. $ 30
C. $ 40
D. $ 46
E. $240

2. Which of the following ordered pairs (x,y) will satisfy the equation $x + y = xy$?

I. (0, 0)
II. (2, 2)
III. (-2,-2)
IV. (2,-2)

F. I only
G. II only
H. I and II only
J. II and III only
K. I, II, III, and IV

3. $(-\frac{1}{2})(-8)(4) = ?$

A. 32
B. 16
C. 0
D. $-4\frac{1}{2}$
E. -16

4. The height of a rectangular solid is 4 meters greater than its width (w). Its length is 2 meters less than 3 times its width. Which expression represents the volume of the solid in cubic meters?

F. $w(3w - 2)(w + 4)$
G. $w(3w - 2)(w - 4)$
H. $w(3w + 2)(w + 4)$
J. $w(3w + 2)(w - 4)$
K. $3w(w - 2)(w + 4)$

GO ON TO THE NEXT PAGE.

5. If the radius of a circle is halved, what happens to the circumference of the new circle?
   - **A.** It remains the same.
   - **B.** It is halved.
   - **C.** It is doubled.
   - **D.** It equals π.
   - **E.** It equals 2π

DO YOUR FIGURING HERE.

6. $\frac{.048 \times 10^5}{2 \times 10^3} = ?$
   - **F.** 1.6
   - **G.** 2.4
   - **H.** 4.8
   - **J.** 1,600,000.0
   - **K.** 2,400,000.0

7. Ann, Bob, Carla, Dave, and Ellie want to play chess. How many different 2-person chess games can be played?
   - **A.** 2
   - **B.** 10
   - **C.** 25
   - **D.** 32
   - **E.** 120

8. Which of the following quantities is NOT equal to $|-7x|$ for all x?
   - **F.** $-7|x|$
   - **G.** $7|-x|$
   - **H.** $7|x|$
   - **J.** $|7x|$
   - **K.** $|-7||-x|$

9. Mary has $\frac{3}{4}$ of her paper route to finish, and John offers to deliver $\frac{1}{3}$ of what remains. What part of the original route does John deliver?
   - **A.** $\frac{1}{4}$
   - **B.** $\frac{1}{12}$
   - **C.** $\frac{1}{8}$
   - **D.** $\frac{1}{2}$
   - **E.** $\frac{7}{12}$

10. The ratio of the measures of the 2 smallest angles of a triangle is 1:2, and the ratio of the smallest to the largest angle is 1:3. What is the measure of the largest angle?
    - **F.** 45°
    - **G.** 77°
    - **H.** 90°
    - **J.** 108°
    - **K.** 135°

GO ON TO THE NEXT PAGE.

2

11. Which equation correctly translates the following statement: "The result of increasing 7 times a quantity by 4 is the same as 12 less than twice the quantity"?

A. $7x + 4 = 2x - 12$
B. $7(x + 4) = 12 - 2x$
C. $7(x + 4) = 12 - (2 + x)$
D. $7x + 4 = 12 - (2 + x)$
E. $7x + 4 = (2 + x) - 12$

12. A square has an area of 64 square centimeters. If each side of the square is increased by 2 centimeters, the total area is increased by how many square centimeters?

F. 4
G. 10
H. 36
J. 44
K. 100

13. On a 30-question test, Toni received 5 points for every correct answer and lost 3 points for every wrong answer. If she answered all questions, how many did she answer correctly if her score was 6 ?

A. 48
B. 25
C. 14
D. 12
E. 11

14. X is a linear function of Y. In converting X to Y, an X score of 5 is equivalent to a Y score of 35, and an X score of 11 is equivalent to a Y score of 65. What Y score is equivalent to an X score of 6 ?

F. 40
G. 45
H. 50
J. 55
K. 60

15. $(3\frac{1}{5})(4\frac{1}{6}) = ?$

A. $7\frac{11}{30}$
B. $12\frac{1}{30}$
C. $12\frac{2}{11}$
D. $12\frac{11}{30}$
E. $13\frac{1}{3}$

16. The system of equations shown below has a solution for which real numbers a and b ?

$$2x + 3y = a$$
$$3x + 8y = b$$

F. Only when $a \neq 0$, $b \neq 0$
G. Only when $a \neq 0$, $b = 0$
H. Only when $a = 0$, $b \neq 0$
J. Only when $a = 0$, $b = 0$
K. For any a and b

DO YOUR FIGURING HERE.

GO ON TO THE NEXT PAGE.

17. If $(x + a)(x + b) = x^2 + 5x + 6$ for all x, then what is the value of ab ?
- **A.** 2
- **B.** 3
- **C.** 5
- **D.** 6
- **E.** 30

DO YOUR FIGURING HERE.

18. Beth wants to cut the largest possible circle from a square piece of material. Which equation expresses the numerical relationship between the side of the material (s) and the radius of the circle (r)?
- **F.** $r = \frac{\sqrt{2}}{2}s$
- **G.** $r = \frac{\sqrt{s}}{2}$
- **H.** $r = \frac{s}{2}$
- **J.** $r = s\sqrt{2}$
- **K.** $r = s$

19. The formula $F = 1.8C + 32$ is used to convert Celsius temperatures to Fahrenheit temperatures. What Celsius temperature is equivalent to 68° Fahrenheit?
- **A.** 180.0°
- **B.** 122.4°
- **C.** 40.0°
- **D.** 20.0°
- **E.** None of the above

20. If $\frac{x - 7}{3} = 2x + 1$, what is the value of x ?
- **F.** 2
- **G.** $\frac{8}{5}$
- **H.** 1
- **J.** $\frac{5}{8}$
- **K.** -2

21. Mr. Stokes waxes [illegible] of a gymnasium floor, and Mr. Wisner waxes [illegible] of the remainder. What part of the floor remains to be waxed?
- **A.** [illegible]
- **B.** [illegible]
- **C.** [illegible]
- **D.** [illegible]
- **E.** [illegible]

22. How is .00634 written in scientific notation?
- **F.** 6.34×10^3
- **G.** $.634 \times 10^2$
- **H.** $.634 \times 10^{-2}$
- **J.** 6.34×10^{-3}
- **K.** 63.4×10^{-5}

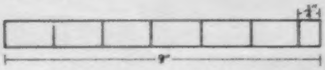
GO ON TO THE NEXT PAGE.

2

DO YOUR FIGURING HERE.

23. Last year, Jane earned $80.00 mowing lawns. This year, she earned $60.00. Compared to her earnings last year, by what percentage did her earnings this year decrease?
    - A. 10
    - B. 20
    - C. 25
    - D. 33
    - E. 75

24. David wants to cut strips of equal width from a board 9 inches long. He makes 6 cuts with a saw that wastes $\frac{1}{8}$ inch of wood with each cut. How many inches wide will the strips be if he ends up with a scrap piece of wood $\frac{1}{4}$ of an inch wide?



    - F. $1\frac{1}{8}$
    - G. $1\frac{1}{4}$
    - H. $1\frac{1}{6}$
    - J. 1 [illegible]
    - K. $1\frac{1}{3}$

25. Point $P(0,4)$ is the center of a circle. Point $A(-2,6)$ is one endpoint of diameter $\overline{AB}$ of this circle. What are the coordinates of point B? (Assume a Cartesian coordinate system.)
    - A. (-4, 8)
    - B. (-2, 2)
    - C. (-1, 5)
    - D. (2, 2)
    - E. (2,14)

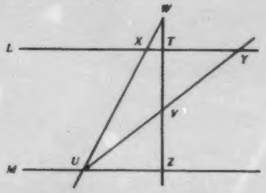
26. If $y = \frac{3x^3}{x-2}$, how much does y decrease as x decreases from 10 to 5 ?
    - F. 5.0
    - G. 12.5
    - H. 18.7
    - J. 25.0
    - K. 225.0

27. There are 3 numbers such that the second is 3 times the first, and the third is 5 more than 4 times the first. If the sum of the numbers is 103, what is the largest number?
    - A. 11
    - B. 12
    - C. $12\frac{1}{4}$
    - D. 43
    - E. 54

GO ON TO THE NEXT PAGE.

DO YOUR FIGURING HERE.

28. In the diagram below, line L is parallel to line M. Of triangles WXT, TVY, WUZ, and ZVU, which pair(s) of triangles is(are) similar?



I. WUZ and WXT
II. TVY and ZVU
III. WUZ and TVY

F. I only
G. II only
H. III only
J. I and II only
K. II and III only

29. Which expressions below are ALWAYS equal when x, y, and z are all positive?

A. $|x + y| - z$ and $|x + y - z|$
B. $|x + y| - z$ and $|x| + |y| - z$
C. $|x + y - z|$ and $x + y + |-z|$
D. $x + |y - z|$ and $x + y + |-z|$
E. $x + y + |-z|$ and $|x| + |y| - z$

30. If dinner rolls cost c cents per roll, what is the cost in dollars of d dozen rolls for any $c > 0$ and for any $d > 0$?

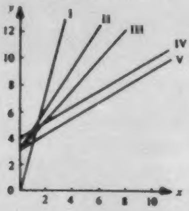
F. $\frac{12cd}{100}$
G. $\frac{12d}{c}$
H. $\frac{12c}{d}$
J. $\frac{cd}{12}$
K. 1,200 cd

31. A wheel rotates 15 times per minute. How many degrees does it rotate in 4 seconds?

A. 60
B. 180
C. 240
D. 360
E. 720

GO ON TO THE NEXT PAGE.

32. Which of the straight lines shown in the figure below is the graph of the equation $y = .5x + 3$?



F. I
G. II
H. III
J. IV
K. V

33. In how many different orders can Joe arrange 5 different books on a shelf?

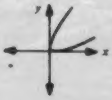
A. 5
B. 10
C. 24
D. 30
E. 120

34. Toni uses 30 feet of fencing to enclose a rectangular garden. If Toni wants the garden to be as large as possible, what dimensions in feet should the garden have?

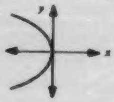
F. 15.0 by 15.0
G. 7.5 by 7.5
H. 5.0 by 10.0
J. 3.0 by 10.0
K. 1.0 by 30.0

35. Which graph shows the equation $x = y^2$?

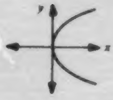
A.



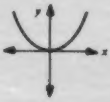
D.



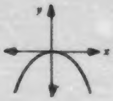
B.



E.



C.



DO YOUR FIGURING HERE.

GO ON TO THE NEXT PAGE.

DO YOUR FIGURING HERE.

36. The diameter of the earth is approximately 4 times the diameter of the moon. The surface area of the moon is approximately what fraction of the surface area of the earth? (Note: The surface area of a sphere of radius r is $4\pi r^2$.)

F. $\frac{1}{2}$
G. $\frac{1}{4}$
H. $\frac{1}{8}$
J. $\frac{1}{16}$
K. $\frac{1}{64}$

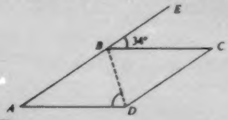
37. What are the values of x that satisfy the inequality $|x - 4| \leq 3$?

A. $x \geq -7$
B. $x \leq -4$
C. $x \leq 7$
D. $x \leq 7 \leq x + 1$
E. $1 \leq x \leq 7$

38. The length of a rectangle is 3 less than twice its width, x. If the area of the rectangle is 35, which equation can be used to find x?

F. $2x + 2(2x - 3) = 35$
G. $2x + 2(2x) - 3 = 35$
H. $x^2 + (2x - 3)^2 = 35^2$
J. $2x^2 - 3x + 35 = 0$
K. $2x^2 - 3x - 35 = 0$

39. A rhombus is a parallelogram with equal sides. In the figure below, $ABCD$ is a rhombus, and the measure of $\angle EBC = 34°$. What is the measure of $\angle BDA$?



A. 17°
B. 68°
C. 73°
D. 79°
E. 124°

40. $\sqrt{65^2 - 56^2} = ?$

F. ± 3
G. ± 11
H. ± 33
J. ± 81
K. ± 121

END OF TEST 2

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 3
SOCIAL STUDIES READING

35 Minutes—52 Questions

DIRECTIONS: Each passage in this test is followed by several questions. After reading a passage, choose the best answer to each question and blacken the corresponding space on your answer sheet. You may refer to the passage as often as necessary.

Adolescence marks the transitional period between childhood and adulthood. There is no set age at which adolescence begins or ends: the first signs of puberty announce its onset, and sexual maturity signals its conclusion. Marked physical, social, and emotional changes occur during adolescence. The onset of puberty is the most striking difference between childhood and adolescence. During puberty, the primary and secondary sex characteristics develop. Primary sex characteristics are the internal and external primary sex glands that carry on reproductive functions. Secondary sex characteristics are the characteristics that distinguish males from females. Other noticeable and rapid physical changes during adolescence affect height, weight, and body form.

Adolescence also involves certain developmental tasks: specific behavioral patterns and skills that society expects individuals to master at certain periods in their life spans. Successful achievement of these tasks leads to happiness and success with later tasks, while lack of success results in unhappiness and failure with later developmental tasks. The prime developmental tasks of adolescence include: (1) achieving emotional independence from parents, (2) accepting one's body and learning to use it effectively, (3) establishing wholesome relationships with age-mates of both sexes, (4) achieving scholastic success, (5) selecting and preparing for an occupation, and (6) achieving economic independence. Compared to the child, the adolescent is less dependent on and less influenced by parents. The adolescent is increasingly engrossed in the larger world outside the home. Peer group acceptance is essential.

Emotionally, adolescence is characterized as a period of stress, storm, and conflict because of physiological changes that create obvious tension. Unlike the child, the adolescent is not in control of his life; he no longer looks like a child but is not ready for adulthood and its demands. Adolescent characteristics are extreme, contradictory tendencies which make this phase of development one of instability and variable moods. Some of these contradictory tendencies are sociability and solitude, independence and dependence, sensitivity and insensitivity, egocentricity and altruism, conformity and nonconformity.

1. Adolescents might typically advise a friend to "join the club unless you want to be left out of everything" primarily because of their:
   - A. natural tendency toward submission.
   - B. previous training in conforming.
   - C. inability to be self-directed.
   - D. need for acceptance and security.

2. Unlike the child's main sources of gratification, the adolescent's stem from:
   - F. peers.
   - G. parents.
   - H. imaginary companions.
   - J. older children.

3. According to the passage, which statement about developmental tasks is most accurate?
   - A. They develop independently of the culture in which one lives.
   - B. They are related primarily to behavioral patterns.
   - C. They are a combination of cultural patterns and skills that develop in stages.
   - D. They are successive but invalid stages of development.

4. All of the following statements characterize the adolescent's peer group EXCEPT that it:
   - F. represents an increasing interest in the outside world.
   - G. provides moral support.
   - H. consumes increasing amounts of time.
   - J. interferes with the adolescent's goals.

5. According to the passage, the search for a satisfying vocational goal can most closely be linked with the adolescent drive toward:
   - A. developing a philosophy of life.
   - B. gaining personal independence.
   - C. achieving scholastic success.
   - D. developing physical competence.

GO ON TO THE NEXT PAGE.

6. According to the passage, which statement about adolescent personality changes is true?

   F. They are influenced by cultural demands.
   G. They remain stable until early adulthood.
   H. They result primarily from physical development.
   J. They result primarily from parental influence.

7. According to the passage, on what levels does adolescent development occur?

   I. Emotional
   II. Physical
   III. Cultural
   IV. Spiritual

   A. I and II only
   B. I, II, and III only
   C. I, II, and IV only
   D. I, II, III, and IV

8. The author suggests that developmental tasks include those skills and goals that:

   F. an individual must possess before reaching maturity.
   G. the social group expects an individual to master.
   H. are acquired entirely as a result of physiological maturation.
   J. are none of the above.

9. Adolescence can be contrasted with childhood upon the appearance of:

   I. primary sex characteristics.
   II. secondary sex characteristics.
   III. emotional stability.
   IV. feelings of dependency.

   A. III only
   B. I and II only
   C. II and IV only
   D. I, II, and III only

GO ON TO THE NEXT PAGE.

3

The increase in black political influence during the 1970s marks a second era of black power in America. By 1900, two black Republicans served in the United States Senate, and twenty black representatives from eight southern states served in Congress. These dramatic electoral breakthroughs of Reconstruction, when former slaves held offices ranging from school superintendent to acting governor, parallel the present success of black candidates in state and local government.

Despite these parallels, the circumstances of political climate, party affiliation, location, and degree of independence differ. During Reconstruction, blacks were elected in a racist climate where many southern whites had been disenfranchised. Today, blacks have been elected in urban centers where both black and white voters go to the polls in free elections. In 1876, blacks faithfully supported the Republican ticket but their descendants have switched to the opposition. During Reconstruction, the rural South was the power base; in this century, northern cities provided the great political opportunities. More important, black constituents influenced passage of the Voting Rights Act of 1965; the Fourteenth and Fifteenth Amendments were the benign work of the white Radical Republicans.

Despite these important differences, some people fear the occurrence of a white backlash like that of 1876. In order to win the presidency, Rutherford B. Hayes agreed to withdraw federal troops from the South. The Compromise of 1876 created innumerable opportunities for disenfranchisement that effectively eliminated minority voters from the rolls and removed black officials from government. With the acquiescence of the Supreme Court, the Fourteenth Amendment guarantees of equal protection became meaningless, and the enactment of "Jim Crow" laws led to the creation of two separate societies.

Strategists today seek to avoid a repetition of that fate by striving for increased black participation to achieve proportionate representation. The Democratic party has been pressed to accept a twenty percent quota for nonwhite convention delegates, and presidential candidates must pledge to reward black supporters with suitable federal appointments. The existence of black mayors at the local level and the Black Caucus in Congress indicates that civil rights protest has matured into the traditional forms of political involvement successfully used by other ethnic and religious minorities. Finally, the major changes that have rendered a very different southland from that of the 1870s encourage the hope that America will live up to its heritage of pluralism and not revert to the pattern of post-Reconstruction reaction.

10. Which activity represents a traditional form of political involvement?

F. NAACP voter registration
G. Black Liberation Army attack on white police
H. Black Panther party community breakfast
J. None of the above

11. Which title best expresses the author's viewpoint?

A. Black Power in Dixie
B. The South Will Rise Again
C. Racist Triumph in U.S. Politics
D. Black Power Revisited

12. The white Radical Republicans would probably have been opposed to the:

F. 15th Amendment.
G. Jim Crow laws.
H. Civil Rights Act of 1964.
J. Voting Rights Act of 1965.

13. If federal registrars were withdrawn from the South, a backlash reaction would be illustrated best by:

A. southern white officials finding ways to reduce the black electorate.
B. black politicans resisting any loss of power.
C. black voters organizing registration drives.
D. the Supreme Court protecting black voting rights.

14. During the post-Reconstruction backlash, southerners would probably NOT have legislated:

F. literacy tests.
G. grandfather clauses.
H. white primaries.
J. compulsory voter-registration laws.

15. The American acceptance of pluralism is best illustrated by the:

A. World War II detention of Japanese-Americans.
B. election of an Irish Catholic to the presidency.
C. Ku Klux Klan's anti-Semitic activities.
D. requirement that all who wish to vote pass tests in English.

16. Which future black strategy would the author probably favor?

F. Mass demonstrations
G. Conventional political campaigns and voter registration
H. Nonviolent protests and civil disobedience
J. Court challenges under the 14th Amendment's equal protection clause

17. According to the first paragraph, which statement best applies to black political power?

A. Certain parallels can be drawn between the post-Civil War era and today.
B. There has recently been an unprecedented growth in black power.
C. National politics has always provided greater opportunities for black advancement than has local government.
D. Limited success at the federal level indicates a need for blacks to focus on local politics.

GO ON TO THE NEXT PAGE.

18. Blacks could request twenty percent of the delegate seats at a Democratic Convention primarily because they:
    - F. live in the most populous states with the largest number of electoral votes.
    - G. have been badly underrepresented in the past.
    - H. are Democrats who hold a high percentage of state and local offices.
    - J. generally contribute a fifth of the vote for Democratic presidential candidates.

19. In the third paragraph, the author implies that the Supreme Court's role concerning black rights during Hayes's administration was to:
    - A. become the early champion of civil rights.
    - B. resolve the dispute over the 1876 election.
    - C. undermine the 14th Amendment.
    - D. insist on equal protection of law, regardless of race.

GO ON TO THE NEXT PAGE.

3

In 1854, when Commodore Matthew C. Perry forced Japan to open its doors to the West, the Land of the Rising Sun was ruled by a military dictator called the *shogun*. Nobles known as *daimyos* employed warriors called *samurai* in their service. The samurai practiced a code of conduct known as *Bushido*, which called for *seppuku*, or ritual suicide, if one were dishonored. The sword was still the standard weapon.

The shogun's inability to resist Perry's demands led to a rebellion, and power passed into the hands of the Emperor Mutsuhito in 1867. Assisted by an informal group of advisors known as the *genro*, Mutsuhito saw that the only way to avoid coming under Western control was to modernize. During his reign, Japan embarked on a program which converted it to a modern nation by 1900.

Taking the advanced nations of the world as their example, the Japanese transformed the samurai into a modern army and navy. These armed forces were tested in the Sino-Japanese War of 1894-95, which established Japanese hegemony over Korea. The triple intervention of Russia, France, and Germany prevented Japan from gaining what it considered to be its just reward for this victory—concessions in southern Manchuria.

Japanese resentment toward Russia, combined with a clash of interests between England and Russia in various parts of the world, led to the Anglo-Japanese Alliance of 1903. This agreement protected Japan against Russia's allies. Japan then provoked and won the Russo-Japanese War of 1904-05.

When Japan established the puppet state of Manchukuo in Manchuria, its conflict with the United States began. Japan's further incursion into China led to war with that nation in 1937; the United States responded with economic sanctions against Japan. On December 7, 1941, Japan reacted by attacking Pearl Harbor, the American naval base in Hawaii. Ultimately, the industrial might of the United States began to prevail, and Japan was frustrated in its efforts to dominate the Far East by force.

20. Manchuria was once described as "the cradle of conflict." What made Manchuria so attractive to Japan and Russia?

- **F.** It was underpopulated and thus attractive to overcrowded Japan.
- **G.** It could provide a shortcut for the Trans-Siberian Railroad to Vladivostok.
- **H.** It was rich in coal and iron, which Japan lacked.
- **J.** All of the above

21. Both the "opening" of Japan and its ultimate defeat in World War II can be attributed in part to the:

- **A.** superiority of the individual Western fighting man.
- **B.** West's advanced technology.
- **C.** Chinese influence on Japanese society.
- **D.** poor choice of allies by Japan.

22. European feudalism differed from Japanese feudalism mainly in the:

- **F.** role of the daimyos.
- **G.** weapons each used.
- **H.** role of the peasantry.
- **J.** practice of seppuku.

23. Other than Thailand, Japan was the only country in the Far East to escape domination by the West during the nineteenth century. According to the author, this is best attributed to Japan's:

- **A.** religious homogeneity.
- **B.** system of alliances.
- **C.** large population.
- **D.** astute leadership.

24. According to the passage, since Japan reacted to American economic sanctions by attacking Pearl Harbor, it is most likely that:

- **F.** imports from Japan were important to the U.S. economy.
- **G.** the sanctions were effective, forcing Japan to react to them.
- **H.** Japan was economically self-sufficient.
- **J.** the U.S. was appeasing Japan.

25. The Anglo-Japanese Alliance of 1903 resulted in part from the clash between England and Russia over such issues as:

- **A.** Russian support for the Afrikaners in the Boer War.
- **B.** the boundary dispute in British Guiana.
- **C.** Russian efforts to oppose the Pan-Slavic movement in the Balkans.
- **D.** Russian expansion in central Asia, which posed a threat to the borders of India.

26. The *genro* was most similar to:

- **F.** the British civil service.
- **G.** Andrew Jackson's Kitchen Cabinet.
- **H.** the U.S. Congress.
- **J.** the secretary of state.

27. The Anglo-Japanese Alliance, which marked an abrupt change in Britain's century-old policy of "Splendid Isolation," most clearly indicated that Britain:

- **A.** recognized the dangers to its interests in the developing system of European alliances.
- **B.** would surrender its claims in China in order to protect its interests in India.
- **C.** was unable to maintain its naval strength.
- **D.** shared the goals of the Open Door Policy.

28. At the beginning of the twentieth century, Japan felt hampered by lack of access to raw materials for its rapidly growing industries and by a shortage of living space for its rapidly growing population. Which other nation believed itself to be in the same position at the same time?

- **F.** Germany
- **G.** Great Britain
- **H.** Russia
- **J.** Spain

GO ON TO THE NEXT PAGE.

Questions 29-43 are not based on a reading passage. You are to answer these questions on the basis of your previous schoolwork in the social studies.

29. The term *cold war* refers to the:
    - A. era of tension between the Soviet Union and the U.S. following World War II.
    - B. struggle between the Soviet Union and China during the 1960s.
    - C. broken pacts of nonaggression signed by the Soviet Union and Germany in the 1930s.
    - D. period of peaceful instability after World War I.

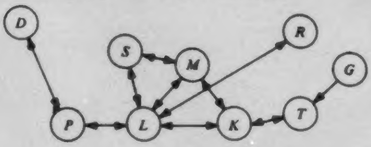
30. In which activity would a young man of ancient Sparta probably have spent the most time?
    - F. Military training
    - G. Practicing the fine arts
    - H. Contemplating religion
    - J. Studying philosophy

31. The United States Senate exercises control over the conduct of foreign affairs primarily because:
    - A. it has the power to veto appropriations voted by the House of Representatives.
    - B. it has the power to confirm treaties.
    - C. the secretary of state must be confirmed by the Senate.
    - D. its relatively small size makes it an appropriate body to advise the president.

32. The Sherman Antitrust Act was intended to:
    - F. prevent consumers from being taken in by fraudulent salespersons.
    - G. break up existing monopolies and prevent the formation of new ones.
    - H. regulate the pricing policies of monopolistic enterprises.
    - J. prevent trust companies from merging with national banks.

33. A substantial increase in the interest rate on consumer loans would result in:
    - A. a rising demand for consumer loans.
    - B. a declining demand for consumer loans.
    - C. a rise in the legal reserve requirement rate on all demand deposits.
    - D. the federal government increasing its borrowing from the public.

34. The results from a questionnaire were used to diagram the sociogram below, which clarifies the:



    - F. group's opinions about social issues.
    - G. social classes the society has established.
    - H. acceptance or rejection patterns among group members.
    - J. group's positions in a formal organization.

35. Panama's objections to the 1904 treaty by which the United States took control of the Panama Canal Zone were based on the contention that:
    - A. President Theodore Roosevelt threatened intervention by U.S. troops if Panama did not accept the treaty.
    - B. Colombia was the rightful owner of the Canal Zone.
    - C. the U.S. has failed to abide by the provisions of the treaty.
    - D. the treaty is an infringement of Panamanian sovereignty.

36. In order to speed up economic recovery in the Soviet Union after World War II, Stalin emphasized:
    - F. quota plans for heavy industry.
    - G. private investment to stimulate industrial activity.
    - H. the production of consumer goods.
    - J. measures to conserve natural resources.

37. Soviet Russia attacked Finland in 1939 mainly because:
    - A. Finland's stable democratic system threatened Soviet Russia.
    - B. Nazi Germany made such an attack the condition for establishing the "Berlin-Moscow axis."
    - C. Soviet Russia was merely trying to regain territory which had been under Russian control during the reign of the czars.
    - D. Finland refused to grant Soviet Russia border territories and rights to forts and naval bases.

38. Which example best supports the notion that "positive traits correlate positively"?
    - F. Drinking too much coffee makes one nervous.
    - G. Identical twins frequently dress identically when they are young.
    - H. One frequently carries an umbrella when dark clouds threaten rain.
    - J. High IQs are usually associated with superior personal adjustment.

GO ON TO THE NEXT PAGE.

39. The method of electing the president by means of an electoral college can be criticized because:

I. states with small populations are over-represented.
II. presidential electors cannot exercise independent judgment since the Constitution binds them to cast their vote for the candidate to whom they are pledged.
III. a candidate might be elected president without getting a majority of the popular votes cast.

A. II only
B. I and II only
C. I and III only
D. I, II, and III

40. During the first century A.D. the Christian Church could be characterized as being:

F. a highly structured organization with widespread support.
G. a popular organization endorsed by the Roman Empire.
H. an underground group of believers supported mainly by the urban lower classes.
J. a hierarchical organization under the pope's rule.

41. An amendment added to a congressional bill that is irrelevant to the bill is called:

A. a filibuster.
B. a rider.
C. an engrossment.
D. a quorum.

42. Regarding their legal basis, local governments in the United States are:

F. legal creations of the U.S. Constitution.
G. completely independent of state governments in almost every state.
H. legal creations of the different states.
J. increasingly independent of both state and federal control.

43. Class mobility in the United States has been aided by the:

I. geographic mobility of the population.
II. availability of plentiful national resources.
III. constitutional elimination of titles of nobility.
IV. policy of universal education.

A. I, II, and III only
B. I, II, and IV only
C. I, III, and IV only
D. I, II, III, and IV

Critics of the petroleum industry charge that it is an oligopoly, dominated by the "majors"—a few huge companies which stifle free competition. They charge that rising gasoline prices and the majors' high profits indicate oligopolistic control.

Most critics argue that this control results from *vertical integration:* engaged in production, refining, distribution, and marketing, the majors control the oil from well to gas pump. This, critics argue, allows the majors to squeeze out smaller companies operating in only one phase and thus to charge artificially high prices. For example, a major could set crude oil prices at a high level, thereby making a large enough profit in production to offset a loss in refining. An unintegrated refining company, however, could not afford to pay the major such high prices for crude oil and would be forced out of business. Finally, some majors produce coal and uranium. Opponents of this *horizontal diversification* argue that this will prevent competition between different energy sources and give the majors control over the entire energy market. These critics urge legislation ordering the majors to operate in only one phase, which would require the divestiture, or selling, of all their assets and facilities in other phases. The legislation would also forbid horizontal diversification.

Industry spokesmen argue that these charges are groundless. They agree that the majors are enormous in size and wealth but argue that size in itself does not produce oligopolistic power. They point out that there are over a hundred companies involved in production, similar numbers in both refining and transport, and several hundred thousand retailers. They also claim that the industry's *concentration ratio*—the percentage of the market controlled by the top four companies—is below thirty-three percent in all phases of the industry, compared to thirty-nine percent for all industries in the United States. These spokesmen also assert that vertical integration provides major efficiencies which ultimately lower the price of their final products. In addition, several federal regulations are supposed to prevent the majors from squeezing out unintegrated companies. They also claim that recent increases in their profit margins are due to a lifting of some governmental controls which had unfairly depressed profits and to the need to finance new exploration efforts. Finally, they argue that oil, coal, and uranium are often found together, and their development requires similar capabilities. Horizontal diversification would therefore lead to lower, not higher, prices for energy.

44. The oil companies claim that some governmental regulations have depressed their profits. Which regulation would be most likely to do so?

F. Placing a ceiling on the price of domestically produced crude oil
G. Allowing immediate deduction of drilling costs in computing taxable income, which would make exploration for oil cheaper
H. Granting permission to build the Alaskan pipeline
J. Granting the depletion allowance, which would reduce the oil companies' taxes

GO ON TO THE NEXT PAGE.

45. Which situation illustrates vertical integration?

- **A.** A salt company sells salt for both eating and industrial purposes.
- **B.** A food processing company grows, packages, and sells its product.
- **C.** A glass company makes windows for both houses and automobiles.
- **D.** None of the above

46. According to the passage, most critics of the oil industry would probably agree that:

- **F.** coal and uranium are better sources of energy than petroleum.
- **G.** increased competition in the oil industry raises gasoline prices.
- **H.** policies aimed at conserving gasoline are futile and counterproductive.
- **J.** the oil companies are gaining control of the entire energy market, given present trends.

47. According to industry spokesmen, horizontal diversification by the major oil companies would result in:

  I. more efficient development of energy resources.
  II. an end to competition between different energy sources.
  III. lower prices for energy.

- **A.** II only
- **B.** III only
- **C.** I and II only
- **D.** I and III only

48. Production of oil and gasoline from deposits within the United States has declined in the last decade. Oil industry spokesmen would probably argue that this is due to all of the following reasons EXCEPT that:

- **F.** federal law requires post-1975 cars to burn unleaded gasoline, which is harder and more expensive to refine.
- **G.** governmental regulations discourage exploration for new deposits.
- **H.** oil companies are holding back production in order to drive up prices.
- **J.** new, offshore oil fields are much harder to develop than those on land.

49. When the author speaks of an "unintegrated" oil company (second paragraph), she means a company which:

- **A.** does not make a big enough profit to stay in business.
- **B.** operates in only one phase of the industry.
- **C.** charges artificially high prices.
- **D.** controls only a tiny percentage of the industry's market.

50. Spokesmen for the oil industry argue that horizontal diversification will:

- **F.** avoid unnecessary and expensive duplication of effort in developing energy sources.
- **G.** counteract the artificially high prices presently being charged by the coal industry.
- **H.** yield more efficiencies than does vertical integration.
- **J.** answer the public's greater demand for oil than for coal and uranium.

51. Spokesmen for the oil industry would probably agree with all of the following statements EXCEPT that:

- **A.** vertical integration is a very efficient form of industrial organization.
- **B.** governmental controls are necessary to protect consumers against excessive fuel prices.
- **C.** the oil industry is more competitive than most American industries.
- **D.** oil companies need a good profit margin to support costly exploration efforts.

52. By "artificially high prices" (second paragraph), the author means prices that are:

- **F.** high enough to produce a large profit for the seller.
- **G.** higher than free market competition would allow.
- **H.** the result of high demand and low supply.
- **J.** the result of low demand and high supply.

END OF TEST 3

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 4

NATURAL SCIENCES READING

35 Minutes—52 Questions

DIRECTIONS: Each passage in this test is followed by several questions. After reading a passage, choose the best answer to each question and blacken the corresponding space on your answer sheet. You may refer to the passage as often as necessary.

A volcano is an opening in the earth's crust through which *magma* (amalgamated volcanic gases and molten rock) is discharged as lava. The volcanic gases that are also emitted are believed to be a combination of water, hydrogen, carbon monoxide and dioxide, nitrogen, argon, sulfur and its oxide compounds, and chlorine.

The intensity of a volcanic explosion depends upon the number and size of the gas bubbles formed in the molten rock beneath the earth's crust. Tiny bubbles of volcanic gas rise within the magma, collide, and merge to form larger bubbles, until a bubble of considerable size and internal pressure forms near the surface of the molten rock. This new bubble expands until the pressure exerted against the wall of the bubble overcomes the surface tension of the molten rock, whereupon the bubble explodes. The bursting of a mass of large, high-pressure bubbles produces a volcanic eruption.

Volcanoes are classified into 6 categories, ranging from least to most intensive: Fissure, Hawaiian, Strombolian, Vulcanian, Plinian, and Pelean.

A Fissure volcano expels large amounts of lava through long horizontal cracks in the earth's surface.

The eruption of a Hawaiian volcano is comparatively quiet and may form lakes of very fluid molten lava.

A Strombolian volcano emits luminous clouds during relatively mild, recurrent eruptions. It shoots ash and hot lava high into the air, where the lava cools into round or spindle shapes before falling to earth.

Between successive eruptions, a Vulcanian volcano forms a plug of hardened lava that is blown out by intense gaseous pressure as large, destructive blocks of rock. It emits a large cloud of gas and ash and may produce a thick, short lava flow or none at all. Volcanic dust permeates the atmosphere and causes changes in the color of the sky.

Gas within a Plinian volcano rushes through great masses of lava, puffing it up and solidifying it into a light and porous rock that contains numerous tiny pockets of gas. Froth and lava flow in glowing rivers down the volcano's exterior.

The Pelean volcano, most explosive and destructive of all, is characterized by the growth of a volcanic dome. It emits incandescent clouds of steam and ash and extremely mobile avalanches of lava that destroy everything in their paths.

1. Which type(s) of volcano would emit small, spindle-like pieces of rock?

   I. Strombolian
   II. Vulcanian
   III. Plinian
   IV. Fissure

   A. I only
   B. I and IV only
   C. II and III only
   D. I, III, and IV only

2. The greatest probable danger at the time of a Fissure volcano's eruption is the:

   F. rapid flow of lava.
   G. resulting layer of volcanic ash.
   H. reaction of hot volcanic gases with cool air.
   J. release of large amounts of volcanic gases.

3. Which meteorological phenomenon most immediately follows a Vulcanian eruption?

   A. Rain, because climatic conditions are disrupted
   B. Cooler temperatures, because new land is formed
   C. Spectacular sunsets, because a large cloud of ash is emitted
   D. Drought, because the lava flows absorb moisture

4. Why is knowledge about the composition of volcanic gases incomplete?

   F. No safe method has been devised for collecting samples of gases from the cone of an erupting volcano.
   G. Methods for studying volcanoes are so new that very few gas samples have been collected.
   H. Methods for separating one gas from another are quite inaccurate.
   J. Some of the gases contained within the magma are not inert.

5. The surface tension of magma allows a gas bubble to escape only when the pressure arising from surface tension:

   A. equals the pressure of the gas bubble.
   B. is exceeded by the pressure of the gas bubble.
   C. is exceeded by the pressure of the molten rock.
   D. is exceeded by the pressure of the surrounding air.

GO ON TO THE NEXT PAGE.

6. Which type of volcano attains a great dome?
   F. Pelean
   G. Strombolian
   H. Vulcanian
   J. Hawaiian

7. Which sequence correctly orders Hawaiian, Plinian, and Vulcanian volcanoes, from the volcano that emits the most gas and viscous lava to the one that emits the least gas and viscous lava?
   A. Hawaiian, Plinian, Vulcanian
   B. Plinian, Vulcanian, Hawaiian
   C. Plinian, Hawaiian, Vulcanian
   D. Vulcanian, Plinian, Hawaiian

8. Large blocks of rock are ejected from Vulcanian volcanoes because gas pressure:
   F. ejects loose rock around the exterior of the volcano.
   G. causes volcanic ash and dust to solidify into large rocks.
   H. disrupts the cone of the volcano, blowing rocks into the air.
   J. expels lava with such force that it is returned to earth as light and porous rock.

9. Why do volcanic gas bubbles rise within the magma under the earth's crust?
   A. Heat within the magma transfers energy to the bubbles, allowing them to rise.
   B. The surface tension of tiny bubbles causes them to rise to the surface of the magma.
   C. Since volcanic gas is lighter than molten rock, the bubbles rise.
   D. The gas bubbles are too large to be constrained by the magma.

GO ON TO THE NEXT PAGE.

4

Nitrogen is a useful and versatile element. Found in many vital substances, nitrogen compounds are important in many biological and chemical reactions.

Atomic nitrogen (N) is never found free in nature, but always in a combined state. As a diatomic gas (N_2), it constitutes the largest fraction of the earth's atmosphere. The N_2 molecule is held together by 3 covalent bonds, each containing a pair of electrons, which make it extremely stable. Diatomic nitrogen can be liquefied at -195.8 degrees Celsius, which makes it a useful coolant in low-temperature experiments.

When nitrogen is combined with oxygen, compounds whose properties depend on the molecular ratio of N to O atoms result. Nitrous oxide (N_2O), commonly called laughing gas, is a colorless, sweet-smelling gas that easily decomposes. Nitric oxide (NO) and nitrogen dioxide (NO_2) are toxic atmospheric pollutants released in the exhausts of combustion engines. Nitrogen dioxide combines with water vapor to form nitric acid (HNO_3), which corrodes the exteriors of buildings and statues. Nitric oxide and nitrogen dioxide are involved in a sequence of reactions that convert ozone in the upper atmosphere to ordinary O_2.

When nitrogen oxides combine with carbon compounds, the resulting compounds can be explosive. Nitroglycerin ($C_3H_5N_3O_9$) was far too difficult to handle safely until Alfred Nobel found a way to absorb it onto finely divided clay; this product is called dynamite. Trinitrotoluene (TNT) is another hydrocarbon compound into which nitrogen oxide groups are incorporated.

Nitrogen is also important in a biological cycle called the nitrogen cycle. A special group of bacteria can "fix" molecular N_2, transforming it into nitrate (NO_3^-), which plants absorb for food. These "nitrogen-fixing" bacteria live in the soil on the roots of certain pealike plants called legumes. Plants transform the nitrates into plant protein, and animals eat the plants and transform plant protein into animal protein. Animal wastes contain urea [$CO(NH_2)_2$] and ammonia (NH_3), which are converted to nitrates through oxidation and bacterial action. Nitrogen-depleted soil can be fertilized with these or similar compounds or rotated in leguminous plants.

10. A particular chemical reaction occurs only after the substances are cooled below -190 degrees Celsius. The coolant used for this experiment is most probably:

F. gaseous N_2O
G. liquid N_2
H. gaseous O_3
J. liquid NH_3

11. A certain compound contains O and N atoms in a 2:1 ratio. Another compound contains these elements in a 5:2 ratio. One conclusion that can be reached about these compounds is that they:

A. are explosive.
B. have different molecular properties.
C. are identical compounds.
D. have similar rates of reaction.

12. Which description defines the properties of a diatomic nitrogen molecule?

F. It is extremely corrosive.
G. It contains a triple bond.
H. It explodes when reacted with carbon dioxide.
J. It quickly reacts to decompose ozone.

13. The historic architecture of many cities is slowly eroding because of industrial air pollution. Which chemical is in part responsible for this corrosion?

A. N_2
B. N_2O
C. NO
D. HNO_3

14. The most likely composition of TNT is:

F. carbon, nitrogen, hydrogen, and oxygen.
G. carbon, nitrogen, and hydrogen.
H. carbon and nitrogen.
J. nitrogen and sulfur.

15. Some persons believe that the operation of supersonic aircraft will deplete the ozone layer, since nitric oxide from exhaust fumes may react with the ozone. This reflects concern about which chemical reaction?

A. $N_2 + 3H_2 \rightarrow 2NH_3$
B. $N_2 + O_2 \rightarrow 2NO$
C. $H_2O + CO(NH_2)_2 \rightarrow CO_2 + 2NH_3$
D. $NO + O_3 \rightarrow O_2 + NO_2$

16. The molecular property known as paramagnetism is generally exhibited by compounds containing unpaired electrons. If C, N, O, and H have 6, 7, 8, and 1 electrons, respectively, which substance is paramagnetic?

F. N_2, which has 14 electrons
G. N_2O, which has 22 electrons
H. NO_2, which has 23 electrons
J. $C_3H_5N_3O_9$, which has 116 electrons

17. Which of the following is NOT an integral part of the nitrogen cycle?

A. Charcoal
B. Air
C. Ammonia
D. Legumes

18. In nitroglycerin ($C_3H_5N_3O_9$), every oxygen atom is bonded to a nitrogen atom, and each carbon atom is bonded to 1 oxygen atom. How many oxygen atoms are NOT bonded to a carbon atom?

F. 0
G. 3
H. 6
J. 9

GO ON TO THE NEXT PAGE.

Questions 19-33 are not based on a reading passage. You are to answer these questions on the basis of your previous schoolwork in the natural sciences.

19. A laboratory experiment requires that a culture have a "minimal medium" furnished for the growth of an organism. This means that the medium should:
    - A. provide optimal growing conditions.
    - B. give supplemental nutrients for growth acceleration.
    - C. contain only those essentials needed for growth.
    - D. contain a chemical which would inhibit reproductive activities.

20. The photosynthetic activity of a plant could be increased by:
    - F. adding magnesium to the soil.
    - G. increasing the humidity in the air.
    - H. reducing the amount of light.
    - J. increasing the amount of light.

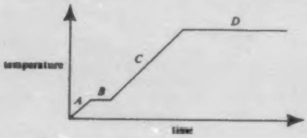
21. Which system(s) could be used as a thermometer?
    - I. A liquid that expands as it is heated in a container
    - II. An electrical resistor that increases its resistance as it is heated
    - III. A junction of dissimilar metals that generates an increasing voltage as it is heated
    - IV. A gas that increases in pressure as it is heated in a closed container

    - A. I only
    - B. II only
    - C. I and II only
    - D. I, II, III, and IV

22. A saltwater solution could be separated into salt and water by means of:
    - F. filtration.
    - G. dilution.
    - H. distillation.
    - J. electrolysis.

23. Which statement explains why, as a person goes up in a balloon, the temperature of the surrounding air decreases?
    - A. The upper air is farther from the hot center of the earth than the lower air.
    - B. There is less heat-giving radioactivity in the upper air.
    - C. The upper air is denser than the surface air.
    - D. None of the above

24. Heat is added at a uniform rate to a beaker of ice that is at a temperature below the freezing point. A record of the temperature of the system is kept, from the time the heating process is started until the subsequent water evaporates. The graph depicts the system's temperature until the melted ice water has evaporated. Why is line *D* longer than line *B*?



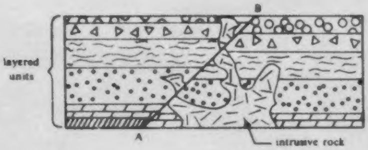
    - F. More heat is needed to change a gram of water to steam than to change a gram of ice to water.
    - G. There is more water to be heated than there had been ice.
    - H. Steam and water cannot exist at the same temperature.
    - J. Steam conducts heat better than water does.

25. The xylem and phloem tissues of vascular plants are analogous to blood tissues in animals because these tissues:
    - A. provide structural support.
    - B. carry dissolved minerals and nutrients.
    - C. supply oxygen to the cells.
    - D. carry carbon dioxide from the cells.

26. When the moisture content of the air is high and temperatures are about 77 degrees Fahrenheit, which situation will most likely occur?
    - F. Cellular division in airborne bacteria will be inhibited.
    - G. Seed plants will become dormant.
    - H. The rate of transpiration in forests will increase.
    - J. Fungi will flourish.

GO ON TO THE NEXT PAGE.

27. In the cross section of rock shown below, the rock layers are cut by an intrusive rock body and by a fault $\overline{AB}$. Which sequence best describes the chronological occurrence of the events depicted, from oldest to youngest?



I. Deposition of layered units
II. Faulting
III. Intrusion

- **A.** I, II, III
- **B.** I, III, II
- **C.** II, III, I
- **D.** III, I, II

28. Which technique should be used to construct a permanent magnet from an unmagnetized iron bar?

- **F.** Bringing the iron bar near a coil through which alternating current is flowing
- **G.** Sawing the ends off the bar to expose its magnetic poles
- **H.** Stroking the bar repeatedly in one direction with one pole of a strong magnet
- **J.** Placing the bar in a strong electric field at a temperature of 0° C

29. Because the specific heat of water is greater than the specific heat of a metal, one knows that water:

- **A.** has a higher specific gravity than most metals.
- **B.** has more mass than most metals.
- **C.** cools more quickly than an equal mass of any metal when the source of heat is removed.
- **D.** needs more heat than an equal mass of metal to experience a rise in temperature equal to that of the metal.

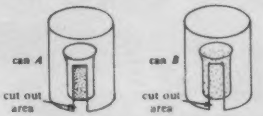
30. Which statement does NOT describe a property of a base?

- **F.** It ionizes to give hydroxide ions.
- **G.** It turns litmus paper blue.
- **H.** It tastes bitter.
- **J.** It reacts with metals to give off hydrogen gas (H_2).

31. The great strength of the carbon-carbon bond in diamonds is a result of which kind of chemical bond?

- **A.** Ionic
- **B.** Metallic
- **C.** Covalent
- **D.** Amorphous

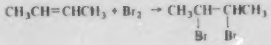
32. In the experiment depicted below, the eyespots of euglenas are removed, and the euglenas are placed in beaker *B* under can *B*. Beaker *A*, which contains normal euglenas, is placed under can *A*. Both cans are then placed in the light. After 24 hours, it is observed that the euglenas in beaker *A* have congregated near the light source, but the euglenas in beaker *B* are evenly dispersed throughout the water.



One of the purposes of this experiment is to determine:

- **F.** the effect of light on euglenas.
- **G.** the function of the eyespots of euglenas.
- **H.** whether euglenas always require light.
- **J.** if euglenas can photosynthesize.

33. Which type of reaction is illustrated by the chemical equation given below?



$$CH_3CH{=}CHCH_3 + Br_2 \rightarrow CH_3\underset{\displaystyle Br}{\underset{|}{CH}}-\underset{\displaystyle Br}{\underset{|}{CH}}CH_3$$

- **A.** Elimination
- **B.** Addition
- **C.** Substitution
- **D.** Replacement

In nuclear fission, the nucleus of an atom is split into 2 pieces. Two forces are involved in this process: the strong *nuclear force*, which is a powerful attraction among neighboring neutrons and protons, and the *electrical force*, which is a repulsion between pairs of protons in the nucleus. The nuclear force is the strongest force that occurs in nature; however, its range is limited so that, like glue, it acts between 2 particles *only* when they are touching each other. On the other hand, the electrical force acts upon *all* protons regardless of the distance between them. Splitting the nucleus into 2 parts leaves the nuclear attraction between the nuclear particles within each half almost unchanged, but permits the electrical repulsive force to separate the parts of the nucleus. The energy of the fission fragments is released as motion.

GO ON TO THE NEXT PAGE.

Fission is possible only for those nuclei that have a large number of protons and, therefore, a large imbalance of electrical energy. But the probability that a nucleus will spontaneously fission is determined by the internal structure of the nucleus. Even though thorium and uranium are the heaviest elements found on earth (that is, they have large numbers of neutrons and protons), they have low probabilities for spontaneous fission. Otherwise their nuclei would have fissioned away (as have much heavier elements) in the 5 billion years since the elements were formed in a stellar explosion.

Adding neutrons to heavy nuclei can induce nuclear fission. However, the process of adding neutrons is effective in inducing fission for only one kind of naturally occurring nucleus: uranium 235. The other naturally occurring elements having heavy nuclei simply "swallow" or absorb neutrons without splitting and form different nuclei.

Because every fission event in an induced nuclear fission reaction releases one or more neutrons, secondary neutrons can also induce fission, if additional fissionable material is present. Thus, a *chain reaction* can be produced. In a *nuclear reactor*, the chain reaction is controlled so that precisely one neutron from each fission event results in one other fission event. If it is not controlled, the reaction will proceed explosively.

34. Possible sources of danger in the operation of nuclear reactors include:
    - F. the release of fission products.
    - G. an uncontrolled chain reaction.
    - H. excess heat.
    - J. all of the above.

35. If nuclear fission did not produce secondary neutrons, which of the following would NOT occur?
    - A. A chain reaction
    - B. Nuclear fission
    - C. An electrical force between protons
    - D. Natural thorium or uranium

36. The energy released in a nuclear fission reaction has its origin in which force?
    - F. Electrical
    - G. Gravitational
    - H. Frictional
    - J. Magnetic

37. The electrical force between 2 protons that are on opposite sides of an atomic nucleus causes the protons to:
    - A. neutralize each other's charges.
    - B. attract each other.
    - C. repel each other.
    - D. change into neutrons.

38. If an element heavier than uranium were discovered among the energetic particles from outer space, the author would probably conclude that the element came from:
    - F. the fission of an element in the earth's atmosphere.
    - G. debris from a nuclear reaction on earth.
    - H. a recent stellar explosion.
    - J. a comet.

39. Which element has the highest probability of spontaneous fission?
    - A. Hydrogen 2
    - B. Carbon 12
    - C. Oxygen 16
    - D. Americium 243

40. When nuclear fission occurs in a large mass of fissionable material, the energy of the fission fragments is:
    - F. released as motion and dissipated as heat.
    - G. destroyed by fission.
    - H. used to reconstruct the nucleus.
    - J. used to attract the fragments into one mass.

41. The age of the earth can be estimated by determining the relative amounts of uranium 235 and uranium 238 in rock samples. Such an estimate would be INVALID if the samples had:
    - A. undergone chemical analysis.
    - B. contained a large percentage of thorium.
    - C. been irradiated with neutrons.
    - D. been heated to a high temperature.

42. According to the passage, the primary difference between spontaneous fission and induced fission is that spontaneous fission:
    - F. releases a great deal of energy.
    - G. produces a chain reaction.
    - H. does not split the nucleus into 2 parts.
    - J. does not require the addition of neutrons.

GO ON TO THE NEXT PAGE.

4

In the mid-1800s, Henry Bates observed that certain palatable butterflies of the family Pieridae, which closely resembled specimens of the butterfly family Heliconiinae, successfully protected themselves from being eaten by "acting" like the unpalatable and unacceptable Heliconiids. This phenomenon, whereby a harmless and palatable organism adapts to the appearance or behavior of a distasteful species in order to evade predators, has been termed *Batesian mimicry*. Three animals are usually involved in this relationship: the *model*, the animal that is avoided for its unpalatability; the *mimic*, the animal that successfully imitates the warning stimuli of the model; and the *receiver*, the animal that notes and is deceived by these likenesses.

Mimicry is not restricted solely to appearance. The development of mimetic color patterns is often accompanied by other changes. Mimetic resemblances show adaptations involving posture, movement, and sound. These modifications are usually accomplished without any alteration of the organism's anatomy. It is incredible that such striking similarities can be produced in the mimic by different structures and chemicals than those in the model.

Since the mimic's success depends upon the receiver's awareness of the distasteful model, the adaptations that the mimic selects are related to the receiver's learning abilities and sensory capacities. The receiver mistakes the mimic for the model because the receiver's senses are confused. Sufficient protection may be provided even if the resemblance of the mimic to the model is not exact. For instance, in localities where the model is found very rarely, the identifying stimuli of the mimic need not be perfect. The presence of both model and mimic at the same time and in the same area, however, requires the mimic to display a greater degree of similarity to the model than when the model is absent.

Scientists theorize that mimetic resemblances have originated in a manner similar to other types of adaptations, that is, as responses to environmental situations. The probability of mimetic adaptation is influenced by several factors, including the size of the model population, the amount of predation of the mimic, and the number of mimics resembling the same model. Through time, protective characteristics are adapted by the mimic; modifications are continued until the organism achieves a satisfactory likeness to the model.

43. Which organisms would be expected to have evolved the largest numbers and varieties of mimics?

A. Mammals
B. Birds
C. Amphibians
D. Insects

44. The receiver is important in Batesian mimicry because:

F. the receiver protects the model from predation.
G. the receiver propagates the model and the mimetic species.
H. without the receiver the adaptations made by the mimic are unnecessary.
J. without the receiver the life cycle of the mimic is incomplete.

45. As a member of a mimetic relationship, a model can be described as an organism that:

A. escapes predation by some animals because of its unpalatability.
B. looks like another animal because it is a member of the same species.
C. adapts to the appearance of another animal.
D. recognizes the warning characteristics of an animal before eating it.

46. For the purposes of mimicry, which characteristic(s) might an organism eventually adapt?

I. A body pattern
II. A digging behavior
III. A buzzing sound

F. I only
G. III only
H. I and II only
J. I, II, and III

47. If a receiver did not have good eyesight, which characteristic(s) might a mimic and a model share for protection?

I. A very intricate body pattern
II. A spiked appendage under the body
III. A peculiar strong odor

A. I only
B. III only
C. I and II only
D. I, II, and III

48. According to the passage, mimetic species have evolved in order to:

F. camouflage the model from the receiver.
G. increase the amount of unpalatable prey.
H. protect themselves from predators.
J. distract the receiver from the model.

49. Under which circumstance would mimicry of a bird species most likely evolve over a long period of time?

A. A severe drought devastates the foliage of an entire area, thereby eliminating the food supply of certain birds.
B. A forest fire denudes a hillside and permanently destroys the nesting sites of many birds.
C. A chemical is sprayed on certain bushes, and birds that eat the berries of these bushes produce eggs with brittle shells.
D. An earthquake forces predators of certain birds to search for new habitats, and these predators reproduce in great numbers.

50. The receiver perceives the mimic as a model because the receiver:

F. recognizes the warning stimuli of the model.
G. learns that both organisms are palatable.
H. attracts both the mimic and the model.
J. has an acute sense of sight.

GO ON TO THE NEXT PAGE.

51. How does mimicry contribute to and support the theory of evolution?

A. Effective mimetic resemblances are produced and modified by the processes of natural selection.
B. Mimetic species show a change over centuries from a greater to a lesser resemblance to the model species.
C. A mimetic species originally possesses all the resemblances to the model species that it ever will.
D. The resemblance of the mimic to the model affects features that are unrelated to the behavior of the mimic.

52. A mimic does not need to closely resemble an unpalatable model if one species inhabits a different area than the other because the:

F. receiver learns to associate a general pattern with a bad taste.
G. mimic no longer competes with the model for food.
H. mimic is not palatable prey for the receiver.
J. same receiver does not gather food from two different areas.

END OF TEST 4.

STOP! DO NOT RETURN TO ANY OTHER TEST.

Answer Key to Sample Test Booklet

Test 1 - English Usage

1. C	11. C	21. B	31. D	41. A	51. A	61. A	71. C
2. H	12. H	22. F	32. G	42. J	52. H	62. H	72. H
3. B	13. B	23. A	33. B	43. A	53. A	63. B	73. D
4. H	14. F	24. G	34. H	44. H	54. H	64. G	74. J
5. B	15. C	25. A	35. A	45. B	55. B	65. C	75. D
6. J	16. J	26. H	36. G	46. G	56. J	66. J	
7. B	17. C	27. C	37. D	47. D	57. A	67. B	
8. J	18. G	28. G	38. G	48. J	58. F	68. F	
9. D	19. C	29. D	39. D	49. D	59. A	69. B	
10. H	20. J	30. J	40. J	50. J	60. F	70. G	

Test 2 - Mathematics Usage

1. A	11. A	21. C	31. D
2. H	12. H	22. J	32. K
3. B	13. D	23. C	33. E
4. F	14. F	24. G	34. G
5. B	15. E	25. D	35. B
6. G	16. K	26. G	36. J
7. B	17. D	27. E	37. E
8. F	18. H	28. J	38. K
9. A	19. D	29. B	39. C
10. H	20. K	30. F	40. H

Test 3 - Social Studies Reading

1. D	11. D	21. B	31. B	41. B	51. B
2. F	12. G	22. J	32. G	42. H	52. G
3. C	13. A	23. D	33. B	43. D	
4. J	14. J	24. G	34. H	44. F	
5. B	15. B	25. D	35. D	45. B	
6. F	16. G	26. G	36. F	46. J	
7. B	17. A	27. A	37. D	47. D	
8. G	18. J	28. F	38. J	48. H	
9. B	19. C	29. A	39. C	49. B	
10. F	20. J	30. F	40. H	50. F	

Test 4 - Natural Sciences Reading

1. A	11. B	21. D	31. C	41. C	51. A
2. F	12. G	22. H	32. G	42. J	52. F
3. C	13. D	23. D	33. B	43. D	
4. F	14. F	24. F	34. J	44. H	
5. B	15. D	25. B	35. A	45. A	
6. F	16. H	26. J	36. F	46. J	
7. B	17. A	27. B	37. C	47. B	
8. H	18. H	28. H	38. H	48. H	
9. C	19. C	29. D	39. D	49. D	
10. G	20. J	30. J	40. F	50. F	

Actual Test Directions

This booklet contains four tests—*English Usage, Mathematics Usage, Social Studies Reading*, and *Natural Sciences Reading*. The four tests measure skills and abilities that are highly related to success in college.

During the first half of the testing session, you will complete the English Usage Test and the Mathematics Usage Test. In the second half, you will complete the Social Studies Reading Test and the Natural Sciences Reading Test.

The questions in each test are numbered, and the four or five suggested answers for each question are lettered. On the answer sheet, the rows of ovals are numbered to match the questions, and the ovals in each row are lettered to correspond to the suggested answers.

For each question, first decide which answer is best. Next, locate on the answer sheet the row of ovals numbered the same as the question. Then, locate the oval in that row lettered the same as your answer. Finally, blacken the oval completely. Use a soft lead pencil and make your marks heavy and black. *DO NOT USE A BALL-POINT PEN.*

You may work on each test ONLY when your test administrator tells you to do so. If you finish a test before time is called, you should use the time remaining to reconsider questions you are uncertain about in that test. You may NOT look back to a test on which time has already been called, and you may NOT go ahead to another test. To do so will disqualify you from the examination.

Your score on each test will be based only on the number of questions you answer correctly. You will NOT be penalized for guessing. *HENCE IT IS TO YOUR ADVANTAGE TO ANSWER EVERY QUESTION.*

If you change your mind about an answer, erase your first mark thoroughly before marking your new answer. For each question, make certain that you mark in the row of ovals with the same number as the question.

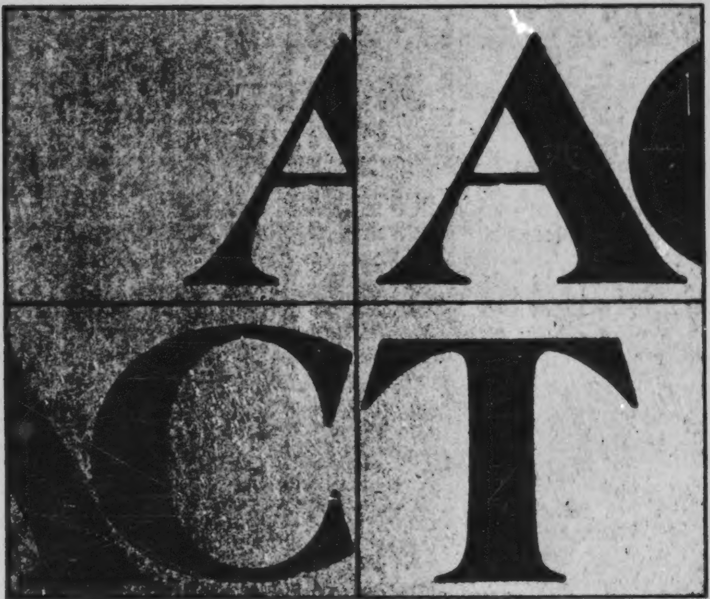
Do not fold or tear the pages of your test booklet. Keep the booklet in as good a condition as possible.

DO NOT OPEN THIS BOOKLET UNTIL TOLD TO DO SO.

The American College Testing Program

The ACT Assessment

Sample Test Booklet*



©1981 by The American College Testing Program. All rights reserved.
*Answer key for this sample test booklet is provided on page 35.

8121C

TEST 1
ENGLISH USAGE

40 Minutes—75 Questions

DIRECTIONS: In the passages that follow, certain words and phrases are underlined and numbered. In the right-hand column you will find alternatives for each underlined part. You are to choose the one that best expresses the idea, makes the statement appropriate for standard written English, or is worded most consistently with the style and tone of the passage as a whole. Choose the alternative you consider best and blacken the corresponding space on your answer sheet. If you think the original version is best, choose "NO CHANGE." Read each passage through once before you begin to answer the questions that accompany it. You cannot determine most answers without reading several sentences beyond the phrase in question. Be sure that you have read far enough ahead each time you choose an alternative.

Passage I

Most people associate the name, Count Dracula[1] with a legendary vampire. They also think that Transylvania, the site of the Counts bloodthirsty deeds,[2] is just as much a fiction as[3] the Count himself. However, this view would be[4] mistaken: Dracula really did live in Transylvania, a part of Rumania that is near the Russian border.[5] In fact, his castle still stands, perched high on a rock, in the village of Bran.

Dracula may never have been a vampire, but he was a ruthless tyrant. He had thousands of his subjects slaughtered and their corpses displayed at the border to scare off an invasion planned by the Turks (the Turks practiced fire worship before converting to Islam)[6] in the fifteenth century.

If Dracula were to return to his old haunts, he would find the customs

1. A. NO CHANGE
   B. name Count Dracula
   C. name, of Count Dracula,
   D. name Count Dracula,

2. F. NO CHANGE
   G. Counts' bloodthirsty deeds,
   H. Count's bloodthirsty deeds
   J. Count's bloodthirsty deeds,

3. A. NO CHANGE
   B. a fiction, like
   C. of a fiction like
   D. fiction, just as

4. F. NO CHANGE
   G. would have been
   H. is
   J. has been

5. A. NO CHANGE
   B. in Transylvania, near the Russian border, a part of Rumania.
   C. near the Russian border, in Transylvania, in a part of Rumania.
   D. a part that is near the Russian border of Rumania in Transylvania.

6. F. NO CHANGE
   G. (before they were converted to Islam, the Turks were fire worshippers)
   H. (before the Turks converted to Islam, they worshipped fire)
   J. OMIT

GO ON TO THE NEXT PAGE.

little changed. Journeys are made frequently by ox-cart, occasionally by bicycle, and by car.[7] Typically, people still cluster in villages and shop in a central marketplace. Chatting with neighbors and watching the passersby are[8] what they enjoy doing. The constant presence of soldiers, marching or stopping pedestrians, creates a depressing ominous[9] atmosphere that Count Dracula would undoubtedly have found to his liking.

7. A. NO CHANGE
   B. only rarely by car.
   C. journeys by car only rarely.
   D. journeys are made only rarely by car.

8. F. NO CHANGE
   G. have been
   H. were
   J. has become

9. A. NO CHANGE
   B. depressing (and ominous)
   C. depressing, and ominous
   D. depressing and ominous

Passage II

Sojourner Truth and Harriet Tubman;[10] two nineteenth-century black Americans, have overcame[11] the physical and psychological bondage of slavery and became vocal champions of freedom. These women contributed significantly to the antislavery movement, the Civil War effort, and the womens rights[12] struggle.

After having gained their freedom—[13] Sojourner Truth by means of legislation and Harriet Tubman by escaping—both women worked for the abolitionist movement. Their deeply, religious convictions[14] led them to

10. F. NO CHANGE
    G. Tubman
    H. Tubman,
    J. Tubman—

11. A. NO CHANGE
    B. have overcome
    C. overcome
    D. overcame

12. F. NO CHANGE
    G. women's rights
    H. women's right's
    J. womens rights'

13. A. NO CHANGE
    B. (Begin new paragraph) After when they gained their freedom—
    C. (Begin new paragraph) After the time their freedom was gained—
    D. (Do NOT begin new paragraph) When after their freedom was gained—

14. F. NO CHANGE
    G. deep religious convictions
    H. deep religious convictions,
    J. deeply religious convictions,

GO ON TO THE NEXT PAGE.

1

an acute awareness and sensitivity to 15 the cause of human liberation. Although uneducated, she made many powerful addresses 16 at antislavery meetings. Harriet Tubman, too, was persuasive as an antislavery orator and conductor on the Underground Railroad; 17 an illegal system of helping slaves escape. Called the "Moses of the Railroad," various techniques were used by Tubman 18 to persuade slaves to make the perilous journey.

Both women contributed to the Civil War effort. Sojourner Truth lectured to the soldiers and collected gifts for them. Harriet Tubman's role, on the other hand, 19 was a more active one. At various times, Tubman served as a cook, a spy, a nurse, a scout, and a guerrilla fighter. Last and finally, these women 20 recognized the similarities between the predicaments of the black slaves, and of women 21 in the nineteenth century. As black females, they possessed a double vision that gave them special insight to 22 the problems of both groups.

Since 23 they made speeches in support of rights for women. Their inspiring words are often cited by modern feminists.

15. A. NO CHANGE
    B. of and sensitivity to
    C. to and sensitivity for
    D. and sensitivity for

16. F. NO CHANGE
    G. many powerful addresses have been made by Sojourner Truth
    H. Sojourner Truth made many powerful addresses
    J. many powerful addresses were made by Sojourner Truth

17. A. NO CHANGE
    B. Railroad,
    C. Railroad; as
    D. Railroad, as

18. F. NO CHANGE
    G. Tubman's various techniques
    H. Tubman, using various techniques,
    J. Tubman used various techniques

19. A. NO CHANGE
    B. therefore,
    C. consequently,
    D. nevertheless,

20. F. NO CHANGE
    G. (Begin new paragraph) These women finally
    H. (Begin new paragraph) Last and finally, these women
    J. (Begin new paragraph) Finally, these women

21. A. NO CHANGE
    B. slaves and of women
    C. slaves and of women,
    D. slaves (and of women)

22. F. NO CHANGE
    G. upon
    H. into
    J. within

23. A. NO CHANGE
    B. However,
    C. Although
    D. OMIT and begin sentence with *They*

GO ON TO THE NEXT PAGE.

Like the North Star that led Harriet Tubman in her escape to freedom, the lives and works of these two women serves as[24] beacons to guide all who fight for human freedom and equality.

24. F. NO CHANGE
G. serve similar to
H. serve just as
J. serve as

Passage III

An ideal, and perfect[25] place for a short vacation, the Black Hills area of South Dakota promises attractive experiences to the outdoorsman, to the history enthusiast, and to the person who enjoys cultural activities.

The outdoor-type person[26] who visits the Black Hills discovers glistening lakes to skim by boat or water skis; in addition, he[27] finds pine-scented mountain trails where you may observe[28] such rare forms of wildlife, as[29] buffalo and antelope. After savoring a fresh trout dinner, he may sleep peacefully under a starlit sky.

The history enthusiast relives the past in Deadwood, where legendary frontier personalities live again during the "Days of '76" celebration.

25. A. NO CHANGE
B. A perfectly, ideal
C. An ideal perfect
D. An ideal

26. F. NO CHANGE
G. outdoors people
H. outdoorsman
J. outdoor-type people

27. A. NO CHANGE
B. skis, in addition;
C. skis furthermore,
D. skis, furthermore,

28. F. NO CHANGE
G. you might observe
H. he may observe
J. a person can glimpse

29. A. NO CHANGE
B. wildlife like:
C. wildlife, like
D. wildlife as

GO ON TO THE NEXT PAGE.

Among the most famous of these are Calamity Jane and Wild Bill Hickok, Jack McCall murdered him as a result of a card game, in Deadwood. He can also visit Bear Butte, where Indian tribes once held council. Finally, where the tourist appreciates cultural activities, these are special: hunting elks, the Black Hills Playhouse productions and the Passion Play. The vast outdoor amphitheater used for these performances is itself an extraordinary sight.

30. F. NO CHANGE
    G. famous ones of these
    H. two most famous of these
    J. very most famous of them

31. A. NO CHANGE
    B. whom Jack McCall during a card game slew
    C. who was murdered by Jack McCall during a card game
    D. who, during a card game, Jack McCall murdered,

32. F. NO CHANGE
    G. (Do NOT begin new paragraph) Finally, when
    H. (Begin new paragraph) Finally, where
    J. (Begin new paragraph) Finally, if

33. A. NO CHANGE
    B. to hunt elk
    C. the hunting of the elks,
    D. OMIT

Passage IV

Having been puzzled over old handwritten manuscripts, we wonder at the ornateness, the elongated letters, and the illegibility, but forget that each writer was following the fashionable style of his time. Those styles, identified by names, like Chancery Cursive, Gothic, and Uncial, reveal eras more than individuals. It is no coincidence that the signatures on the Declaration of Independence strike us as so similar: for the Founding Fathers, fashion not only had shaped dress and architecture, but also handwriting.

34. F. NO CHANGE
    G. Having been puzzled over old, and
    H. Puzzling over old,
    J. Puzzled over old, and

35. A. NO CHANGE
    B. names, such as
    C. names such as
    D. such names, as

36. F. NO CHANGE
    G. (Place after *Fathers.*)
    H. (Place after *had*)
    J. (Place after *shaped*)

GO ON TO THE NEXT PAGE.

Only yesterday our grandparents carefully copied the flowing Palmer or Spencer scripts. Today we may have learned Manuscript or Cursive, yet we no longer follow any [37] common style. In an age of international rock and roll, unisex dress, and the common denominator of television, which seems strange, that a [38] universal handwriting style does not reemerge; however, [39] our age is dominated by many communication devices that create lesser need [40] for the handwritten word. Perhaps this technology makes us indifferent with [41] the dictates of fashion. Conversely, perhaps our individualized scripts allow us to be attractive, fanciful, practical—even illegible—at whatever slant you choose. [42]

37. A. NO CHANGE
    B. each
    C. any form of
    D. any type of

38. F. NO CHANGE
    G. (which seems strange)
    H. it seems strange, that
    J. it seems strange that

39. A. NO CHANGE
    B. moreover,
    C. therefore,
    D. thus,

40. F. NO CHANGE
    G. less needs
    H. less need
    J. fewer need

41. A. NO CHANGE
    B. to
    C. for
    D. from

42. F. NO CHANGE
    G. we choose.
    H. I choose.
    J. is chosen.

GO ON TO THE NEXT PAGE.

1

Passage V

Wilderness areas are a unique part of our national land-preserve system. Different from the national parks and national forests, the Wilderness Act of 1964 established these national wilderness areas. [43]

Areas are selected to [44] inclusion in the wilderness system because each and every one of them [45] have not yet been changed by man. They are places where the earth and its [46] community of life are untrammeled by man. At where [47] man himself is a visitor who [48] does not remain. In such areas, man must travel by foot or on horseback. Because [49] motorized vehicles are not allowed. Nor may buildings be constructed, timber cut, or roads may not be [50] built.

Wilderness areas serve several purposes. It provides [51] a place for wildlife to evolve freely, for plant life to grow naturally, and for people to enjoy the "tonic of wildness."

43. A. NO CHANGE
    B. the government passed the Wilderness Act of 1964.
    C. in 1964 the Wilderness Act was passed by Congress and established these national wilderness areas.
    D. national wilderness areas were established with the passage of the Wilderness Act of 1964.

44. F. NO CHANGE
    G. for
    H. with a view to
    J. by

45. A. NO CHANGE
    B. they
    C. each of them
    D. every one of them

46. F. NO CHANGE
    G. it's
    H. their
    J. our

47. A. NO CHANGE
    B. man, where,
    C. man, to where
    D. man, where

48. F. NO CHANGE
    G. visitor, who
    H. visitor—that—
    J. visitor; who

49. A. NO CHANGE
    B. horseback because
    C. horseback, in that
    D. horseback: because

50. F. NO CHANGE
    G. or roads
    H. and no roads may be
    J. and roads may not be

51. A. NO CHANGE
    B. It offers
    C. It serves as
    D. They provide

GO ON TO THE NEXT PAGE.

Passage VI

Most people would not choose a leopard as a house pet. Yet an American woman, whose husband was stationed in India, for two years 52 did manage to obtain a baby infant 53 leopard soon after its birth and to rear it as a companion for her family.

At first 54 her plan which was to result in bringing 55 one of the most savage of the big cats into her household met with opposition from Indian officials, friends, and neighbors worried about her safety. But she overcame their protests and settled the leopard in her home when it was only twenty-one days old.

As Cat (which was its name) grew, its energy and size were the cause of constant household disorder. And 56 in spite of mounting wreckage, the family grew to love their strange pet.

Before the family returned to the States, their 57 Cat was placed in the Delhi Zoo. After 58 a heartbreaking period of weaning it from constant human companionship to the solitary life 59 of an animal behind bars.

52. **F.** NO CHANGE
    G. India for two years,
    H. India (for two years)
    J. India, for two years,

53. **A.** NO CHANGE
    B. baby, infant
    C. baby and infant
    D. OMIT

54. **F.** NO CHANGE
    G. First,
    H. First off
    J. In the first place,

55. **A.** NO CHANGE
    B. brought
    C. resulting in the bringing of
    D. to bring

56. **F.** NO CHANGE
    G. Thus,
    H. But
    J. Although

57. **A.** NO CHANGE
    B. the
    C. her
    D. OMIT

58. **F.** NO CHANGE
    G. Zoo. It was after
    H. Zoo. But after
    J. Zoo—but only after

59. **A.** NO CHANGE
    B. life in solitude
    C. life all alone
    D. lone life

GO ON TO THE NEXT PAGE.

1

Passage VII

> These paragraphs may or may not be in the most logical order. The last item will ask you to choose the most logical order.

(1)

Feminism did not die, but a depression, a war, and the Freudian belief in female inferiority effectively thwarted much activism. Daughters were content to consolidate the gains privately their mothers had won publicly. In the mid-sixties, however, feminism reemerged again. Although the feminist movement might strike those under twenty-five as novel and revolutionary, there is ample evidence, that women have worked for over three hundred years to obtain equal rights.

(2)

Although widely used to identify the women's movement, the term, *feminism* is of rather current coinage. Not until the 1890s did it assume the meaning we now accept. Sexual equality. Within today's political context, it has been comfortably assimilated into our language.

(3)

That contemporary historians have had to use the term demonstrates that history has traditionally avoided focusing on women. Because our great historians

60. F. NO CHANGE
    G. privately (Place after *Daughters*)
    H. in private (Place after *were*)
    J. privately (Place after *consolidate*)

61. A. NO CHANGE
    B. once again.
    C. yet another time.
    D. OMIT and end sentence with *reemerged*.

62. F. NO CHANGE
    G. evidence; that,
    H. evidence that
    J. evidence that,

63. A. NO CHANGE
    B. term, known as *feminism*,
    C. term *feminism*
    D. term *feminism*—

64. F. NO CHANGE
    G. accept: sexual
    H. have excepted; that being sexual
    J. have accepted. That of sexual

65. A. NO CHANGE
    B. in
    C. within
    D. through

GO ON TO THE NEXT PAGE.

have been males, there [66] has been an unconscious, sexist, interpretation. [67] Consequently, women have remained invisible in textbooks and records, the [68] assumption being that they have made negligible contributions. Feminism, moreover, [69] was quite visible in colonial America when women petitioned for rights like those enjoyed by men. Later, during the Revolution, Abigail Adams implored her husband John to remember the men [70] and be generous else they'd rebel. Adams, however, affects [71] no female independency.

(4)

The crucial year was 1848, with the first women's convention meeting in New York. They [72] succeeded in drawing up eleven resolutions; the [73] most controversial was the demand for suffrage. Their activities suspended during the Civil War [74] women were compelled not only to endure disenfranchisement until 1920, but also to suffer a forty-year hibernation thereafter.

66. F. NO CHANGE
    G. males, there's
    H. males. There's
    J. males, theirs

67. A. NO CHANGE
    B. unconsciously sexist interpretation.
    C. interpretation which is sexist, though unconsciously so.
    D. unconsciously, sexist interpretation.

68. F. NO CHANGE
    G. records the
    H. records. The
    J. OMIT

69. A. NO CHANGE
    B. therefore,
    C. for example,
    D. however,

70. F. NO CHANGE
    G. the rights
    H. the ladies
    J. OMIT

71. A. NO CHANGE
    B. effects
    C. affected
    D. will affect

72. F. NO CHANGE
    G. Women
    H. The delegates
    J. People

73. A. NO CHANGE
    B. resolutions, the
    C. resolutions and the
    D. resolutions but

74. F. NO CHANGE
    G. War. However
    H. War,
    J. War, but

75. Choose the sequence of paragraph numbers that will make the essay's structure most logical.
    A. NO CHANGE
    B. 2, 1, 3, 4
    C. 2, 3, 4, 1
    D. 4, 2, 3, 1

END OF TEST 1

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 2
MATHEMATICS USAGE
50 Minutes—40 Questions

DIRECTIONS: Solve each problem, choose the correct answer, and then blacken the corresponding space on your answer sheet.

Do not linger over problems that take too much time. Solve as many as you can; then return to the others in the time you have left.

Note: Illustrative figures are not necessarily drawn to scale and are assumed to lie in a plane unless otherwise stated.

DO YOUR FIGURING HERE.

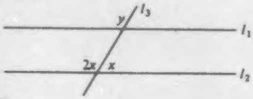
1. The world's largest gem diamond, the Cullinan, weighs approximately 3,100 carats. If 5 carats equal 1 gram, and 31 grams equal 1 ounce, what is the approximate weight of this diamond in ounces?

A. $1\frac{1}{4}$
B. 2
C. 20
D. 50
E. 1,922

2. If each of the following were regular polygons and were inscribed in the same circle, which polygon would have the greatest area?

F. Triangle
G. Square
H. Pentagon
J. Hexagon
K. Octagon

3. In the figure, l_1 and l_2 are parallel lines cut by the transversal l_3. If the measures of the angles are as shown, then y = ?



A. 60°
B. 120°
C. 150°
D. 180°
E. 360°

4. In a certain game, Sue must try to guess the number that Betty has written on a slip of paper. If Sue guesses correctly, she receives 10 points; if she guesses incorrectly, she loses 2 points. After 7 turns, Sue's total score is 22. How many incorrect guesses did she make?

F. 1
G. 3
H. 4
J. 11
K. 14

GO ON TO THE NEXT PAGE.

5. What percent is equivalent to 0.3295 ?

A. 0.003295%
B. 0.03295%
C. 0.3295%
D. 3.295%
E. 32.95%

DO YOUR FIGURING HERE.

6. For all x, $(2x - 3)^2 = ?$

F. $4x^2 - 6x - 6$
G. $4x^2 - 6x - 9$
H. $4x^2 - 12x + 6$
J. $4x^2 - 12x + 9$
K. $4x^2 + 12x + 9$

7. $\frac{3}{24} \div (-\frac{2}{3}) = ?$

A. $-\frac{16}{3}$
B. $-\frac{3}{16}$
C. $-\frac{1}{12}$
D. $\frac{1}{12}$
E. $\frac{3}{16}$

8. The formula $F = \frac{9}{5}C + 32$ can be used to convert Celsius (C) to Fahrenheit (F) temperature. To the nearest tenth, what is the Celsius equivalent of 60° Fahrenheit?

F. 15.6°
G. 20.0°
H. 29.7°
J. 29.8°
K. 140.0°

9. The rate of a long-distance telephone call between 2 cities is $1.50 for the first 3 minutes plus $.35 for each additional minute. If a long-distance call cost $3.25, then how many minutes long was the call?

A. 3
B. 5
C. 6
D. 8
E. 9

10. For all x, $(x^3 + 2x^2 + 3x - 6) + (5x^3 - 2x^2 + 6) + (-6x^3 + 5x^2 - 18x) = ?$

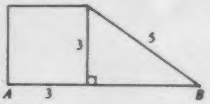
F. $5x(x - 3)$
G. $x(5x + 21)$
H. $3x(3x - 5)$
J. $x(12x^2 + 5x - 15)$
K. $5x^2 - 15x + 12$

GO ON TO THE NEXT PAGE.

2

11. What is the length of $\overline{AB}$ in the figure shown below?

DO YOUR FIGURING HERE.



A. 6
B. 7
C. 8
D. 9
E. 10

12. A man bought a new car for $6,000. He arranged to pay $\frac{1}{3}$ of the total price at the time of the sale, $\frac{1}{2}$ of the balance in 30 days, and the remaining balance at the end of 60 days. How much was due at the end of 60 days?

F. $1,000
G. $1,500
H. $2,000
J. $3,000
K. $4,000

13. Which of the following would be a correct application of the distributive property of multiplication over addition?

A. $2(3 + 4) = (2 + 3)(2 + 4)$
B. $2(3 + 4) = (2 \cdot 3) + (2 \cdot 4)$
C. $2 + (3 \cdot 4) = (2 + 3)(2 + 4)$
D. $2 + (3 \cdot 4) = (2 \cdot 3) + (2 \cdot 4)$
E. All of the above

14. Suppose $\frac{1}{4}$ of a worker's salary is deducted for income taxes, and $\frac{3}{5}$ of income taxes are used for military expenditures. If $3,000 of a worker's pay were used for military expenditures, then which equation could be solved to determine the worker's salary?

F. $(\frac{1}{4}x)(\frac{3}{5}) = 1,800$
G. $(\frac{1}{4}x)(\frac{3}{5}) = 3,000$
H. $\frac{1}{4}x = 1,800$
J. $\frac{1}{4}x + \frac{3}{5}x = 3,000$
K. $\frac{1}{4}x - \frac{3}{5}x = 3,000$

15. The sum of $\frac{5}{8}$ of a certain number, x, and $\frac{1}{2}$ of the same number is 18. Which of the following equations could be used to solve for x?

A. $\frac{5}{8}x + \frac{1}{2} = 18$
B. $x + \frac{5}{8} + \frac{1}{2} = 18$
C. $\frac{5}{8}x + \frac{5}{16}x = 18$
D. $\frac{5}{8} + \frac{1}{2}x = 18$
E. $\frac{5}{8}x + \frac{x}{2} = 18$

GO ON TO THE NEXT PAGE.

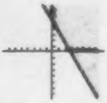
16. A circle is centered at the origin of a Cartesian coordinate system. If the coordinates of one endpoint of a diameter of the circle are $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$, then what are the coordinates of the diameter's other endpoint?
    - F. $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$
    - G. $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$
    - H. $(-\frac{\sqrt{3}}{2}, -\frac{1}{2})$
    - J. $(\frac{\sqrt{3}}{2}, -\frac{1}{2})$
    - K. $(\frac{1}{2}, -\frac{\sqrt{3}}{2})$

17. A rectangle is 12 meters long and 5 meters wide. If the length and the width of the rectangle are increased by the same amount, by how many meters should each dimension be increased to double the area of the rectangle?
    - A. 1
    - B. 3
    - C. 4
    - D. 15
    - E. 20

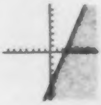
18. $\{x \mid x \text{ is a real number and } \sqrt{x} = 25\} = ?$
    - F. $\{625\}$
    - G. $\{5\}$
    - H. $\{-5\}$
    - J. $\{-625\}$
    - K. ϕ

19. Which graph correctly illustrates the relation $y \leq 5 - 2x$?

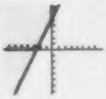
A.



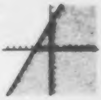
D.



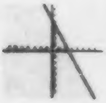
B.



E.



C.



DO YOUR FIGURING HERE.

GO ON TO THE NEXT PAGE.

20. An aquarium is 15 inches long, 11 inches wide, and 10 inches high. How many gallons of water are needed to fill it to 3 inches below the top if 1 gallon occupies 231 cubic inches?

F. 5.0
G. 7.5
H. 8.0
J. 10.0
K. 19.5

21. If $9x^2 + 6x(x - 4) - 180 = 6x + 45$, then $x = ?$

A. -5 or 3

B. 5 or -3

C. $\frac{3 \pm \sqrt{234}}{5}$

D. $\frac{-1 \pm i\sqrt{14}}{3}$

E. $\frac{6 \pm 16\sqrt{6}}{5}$

22. Which of the following is the largest number?

F. 5.0×10^{-7}
G. 2.0×10^{-7}
H. 2.1×10^{-6}
J. 5.0×10^{-8}
K. 1.80×10^{-6}

23. If $3,864 was spent for a car which was discounted to 84 percent of the original price, what was the original price?

A. $2,100.00
B. $3,245.76
C. $4,482.24
D. $4,600.00
E. $7,109.76

24. What positive value of x satisfies the equation $9^{x^2} = 81^2$?

F. $\sqrt{2}$
G. 2
H. 3
J. 4
K. 9

25. A dealer buys gasoline from his wholesaler for 50¢ per gallon. He must pay the following taxes on this cost:

17.2% federal highway tax
22.8% state highway tax
5% state sales tax

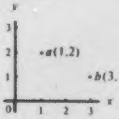
At what price (in cents per gallon) must he sell the gasoline in order to make 2.1¢ per gallon profit?

A. 97.1
B. 74.6
C. 72.5
D. 72.1
E. 56.6

DO YOUR FIGURING HERE.

GO ON TO THE NEXT PAGE.

26. What is the distance between points a and b in the Cartesian coordinate system shown below?



F. $\sqrt{3}$
G. $\sqrt{5}$
H. $2\sqrt{2}$
J. 3
K. 5

DO YOUR FIGURING HERE.

27. If Mary rides her bicycle east at the rate of 16 kilometers per hour and returns over the same route at the rate of 24 kilometers per hour, how many kilometers east can she travel if she bicycles for a total of 5 hours?

A. 8
B. 20
C. 40
D. 48
E. 100

28. The recipe for 18 cookies calls for $\frac{2}{3}$ cup sugar. How many cups of sugar will be needed to make 4 dozen cookies?

F. $1\frac{7}{9}$
G. $2\frac{2}{3}$
H. 3
J. $3\frac{5}{9}$
K. 4

29. $|x - 3| < 9$ if and only if which of the following is true?

A. $x > -6$
B. $x < 12$
C. $-6 < x < 12$
D. $x < -6$ or $x > 12$
E. $x > -6$ or $x > 12$

30. If $g(x) = 2x + 1$ and $f[g(x)] = 7x - 1$, then $f(2) = $?

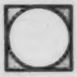
F. $\frac{5}{2}$
G. 5
H. $\frac{13}{2}$
J. 13
K. 34

GO ON TO THE NEXT PAGE.

31. If a side of an equilateral triangle is 8 meters in length, what is the length, in meters, of its altitude?
   - **A.** $2\sqrt{3}$
   - **B.** 4
   - **C.** $2\sqrt{6}$
   - **D.** $4\sqrt{3}$
   - **E.** 16

DO YOUR FIGURING HERE.

32. The circle inscribed in the square below has a radius of 2 inches. What is the area, in square inches, of the shaded area?



   - **F.** $4 - 4\pi$
   - **G.** $8 - 4\pi$
   - **H.** $4(4 - \pi)$
   - **J.** $2(8 - \pi)$
   - **K.** $8 - 8\pi$

33. Which of the following is true if set $A = \{1,2,3,4,5,6,7,8,9,10\}$; $B = \phi$, $C = \{5,6,7\}$, and $D = \{0,1,2,3\}$?
   - **A.** $(C \cup D) \subset A$
   - **B.** $(B \subset A)$ and $(C \subset A)$
   - **C.** $(D \subset A)$ and $(C \subset A)$
   - **D.** $(B \subset D)$ and $(D \subset A)$
   - **E.** $(D \subset A)$ and $(B \subset A)$

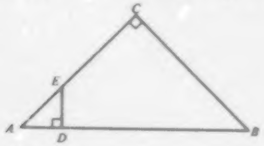
34. How many different 7-digit telephone numbers (for example, 923-5678) can be written if the digit 0 cannot appear among the first 3 digits?
   - **F.** $2^9 \times 2^8 \times 2^7 \times 2^{10} \times 2^9 \times 2^8 \times 2^7$
   - **G.** $9 \times 8 \times 7 \times 7 \times 6 \times 5 \times 4$
   - **H.** $2^9 \times 2^9 \times 2^9 \times 2^{10} \times 2^{10} \times 2^{10} \times 2^{10}$
   - **J.** $9 \times 9 \times 9 \times 10 \times 10 \times 10 \times 10$
   - **K.** $9 \times 8 \times 7 \times 10 \times 9 \times 8 \times 7$

35. For a given fraction, $\frac{a}{b}$, the numerator is halved and the denominator is doubled. The resulting fraction could have been obtained by multiplying the original fraction by which quantity? ($a \neq 0$; $b \neq 0$)
   - **A.** $\frac{1}{4}$
   - **B.** $\frac{1}{2}$
   - **C.** 1
   - **D.** 2
   - **E.** 4

GO ON TO THE NEXT PAGE.

DO YOUR FIGURING HERE.

36. In $\triangle ABC$, $\overline{AC} \perp \overline{BC}$ and $\overline{ED} \perp \overline{AB}$. If the length of $\overline{AD} = 4$, $\overline{AE} = 6$, and $\overline{AB} = 24$, then what is the length of $\overline{AC}$?



- **F.** 8
- **G.** 12
- **H.** 16
- **J.** 18
- **K.** 20

37. If $\log_{10} 2 = .3010$ and $\log_{10} 7 = .8451$, $\log_{10} 98 = ?$
- **A.** .2150
- **B.** .5087
- **C.** 1.0152
- **D.** 1.9912
- **E.** 2.9912

38. $\{x \mid 6x^2 - 7x - 3 = 0\} = ?$
- **F.** $\{\frac{1}{3}, \frac{2}{3}\}$
- **G.** $\{-\frac{1}{3}, \frac{3}{2}\}$
- **H.** $\{\frac{1}{3}, -\frac{2}{3}\}$
- **J.** $\{-\frac{1}{3}, -\frac{3}{2}\}$
- **K.** $\{-3, \frac{3}{2}\}$

39. What value of m is required in order that the graph of $y = mx + b$ pass through points $(1,1)$ and $(-1,-5)$?
- **A.** -3
- **B.** -2
- **C.** 2
- **D.** 3
- **E.** 4

40. If $\sqrt{13^2 - 12^2} = \sqrt[n]{125}$, then $n = ?$
- **F.** $\frac{1}{5}$
- **G.** $\frac{1}{3}$
- **H.** 2
- **J.** 3
- **K.** 5

END OF TEST 2

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 3

SOCIAL STUDIES READING

35 Minutes—52 Questions

DIRECTIONS: Each passage in this test is followed by several questions. After reading a passage, choose the best answer to each question and blacken the corresponding space on your answer sheet. You may refer to the passage as often as necessary.

The peasant farmers of Southeast Asia, who have stoutly resisted planting higher-yielding varieties of rice or new crops that could be sold for large profits, have baffled Western economists. The peasants do not seem to want better crops and more income, a situation some have smugly attributed to the "cultural backwardness" of peasant society.

The peasants' actions make sense, however, given the basic fact of peasant life: they live close to the subsistence level—the minimum amount of food needed to survive. If a crop failure occurs, they are in immediate danger of starving. Therefore they prefer traditional methods, which produce only modest yields in good years but are likely to yield something even under poor conditions. Although new varieties of rice produce much more food in the long run, they are more susceptible to drought, flood, or disease. Because the peasants can ill afford additional risks, they stick to traditional methods of cultivation.

This subsistence ethic pervades peasant life. First, peasants will work tremendously hard to produce very small increases in their crop's yield. Second, social status depends on the security of one's income rather than the amount. Although wage laborers generally make more money than farmers, the farmers' position is preferable: in poor times, farmers can at least eat what they have grown. Similarly, peasants prefer to own a little land rather than rent a large plot, even though the latter will yield a larger income in most years.

The Western powers that colonized Southeast Asia were tragically ignorant of the subsistence ethic. Although the traditional kingdoms of the region extracted large portions of the farmers' surpluses in good years for taxes, they collected no taxes in bad years and, if possible, distributed food to the hungry. Taxes were also levied on villages, not individuals. Those whose crops had done better could and would pay more. The colonial powers reversed these policies. Although average taxes were lower, they were the same in bad years as in good, and everyone had to pay the same amount. These policies helped produce a general decline in the peasants' economic well-being, widespread famines, and ultimately, several peasant rebellions.

1. The methods of taxation used by the traditional kingdoms of Southeast Asia insured higher taxes from those peasants who were economically well off. Which form of taxation in the United States is based on this same principle?
   A. Sales taxes on retail goods
   B. Taxes on liquor and cigarettes
   C. Entertainment taxes on nightclubs and discotheques
   D. The progressive federal income tax

2. The peasants resisted planting cash crops, which they could have sold for large profits, mainly because they:
   F. lacked the knowledge needed to grow the cash crops.
   G. would have had to hire wage laborers in order to grow cash crops.
   H. would have had to pay high rents on the additional land needed to grow cash crops.
   J. felt that it was safer to grow crops that could be eaten than crops that had to be sold.

3. Why would Southeast Asian peasants work tremendously hard only "to produce very small increases in their crop's yield"?
   A. For people living near the subsistence level, a little more food is very valuable.
   B. Social status in peasant society depends on the amount of one's income.
   C. The traditional kingdoms of the area place a constant, high tax burden on them.
   D. The peasants fail to realize that this work is very unprofitable.

4. The author's main intention in this passage is probably to:
   F. show that the apparently irrational behavior of the peasants is rational, given the demands of their economic situation.
   G. describe the cultural backwardness of peasant society.
   H. show how the peasants could be helped by applying the knowledge of Western economics.
   J. describe the techniques that succeeded in persuading peasants to adopt new, more efficient methods of cultivation.

GO ON TO THE NEXT PAGE.

5. Which statement best reflects the attitude of the peasants toward taxation?

   A. What matters is not how much is taken for taxes, but how much is left.
   B. A moderate tax that is always the same is preferable to a tax that is sometimes high, sometimes low.
   C. Taxation is fair if everyone pays the same amount.
   D. Taxes should be levied on property, not on income.

6. Western economists have been baffled by the behavior of Southeast Asian peasants primarily because the peasants:

   F. demonstrated cultural backwardness.
   G. seemed to act contrary to their economic self-interest.
   H. demonstrated a lack of concern for social status.
   J. stoutly resisted coming into contact with the economists.

7. The economic behavior of a United States plumbing supply business would be comparable to that of Southeast Asian peasants if it:

   I. borrowed heavily in order to expand rapidly, even though this increased the risk of bankruptcy.
   II. took actions that hurt its profit margin, but helped it retain its present customers.
   III. concentrated on selling products for which there was a steady demand, but which yielded relatively low profits.

   A. II only
   B. III only
   C. I and II only
   D. II and III only

8. Some economists in this country advocate a *negative income tax* (instead of paying taxes, people earning less than a certain figure would receive money from the government). Which of the following policies of traditional Southeast Asian kingdoms would be similar to the negative income tax?

   I. In good crop years, large portions of farmers' surpluses are extracted for taxes.
   II. Taxation is levied on villages instead of individuals.
   III. In bad crop years, no taxes are collected, and food is distributed to the needy.

   F. I only
   G. II only
   H. III only
   J. I and II only

9. Peasant farmers would be more likely to grow the new varieties of rice, which have higher average yields, if their government:

   A. provided them with fertilizers and mechanized farm equipment to do so.
   B. built facilities where they could safely store surplus rice from good years until it was needed in bad years.
   C. instructed them on the proper techniques for cultivating the new varieties.
   D. instituted mass education programs to combat the cultural backwardness of peasant society.

GO ON TO THE NEXT PAGE.

3

From antiquity through medieval times, dreams were considered supernatural and prophetic manifestations. Although modern psychologists have been studying the nature and meaning of dreams, they still have no clear-cut definition of dreaming. However, much new information about the dreaming process is available because of research in sleep laboratories. We now know, for example, that a night's sleep is made up of several cycles of approximately ninety minutes in length. Each cycle is divided into five different levels or stages of sleep. The most vivid dreaming occurs in the last stage of the sleep cycle, which is called *REM* (rapid eye movement) sleep. People awakened during this period of sleep nearly always have clear memories of dreams. Although dreaming occurs during the other stages of sleep, it is less detailed and less frequent. We also know now that dreaming is important to mental health. If people are deprived of REM sleep for several nights, they will begin to hallucinate during waking hours. Finally, sleep laboratory research has also disproved several old myths about dreams. Dreams do not usually take place in an instant. Dreams that incorporate color are not limited to a few people. No one is dreamless.

Modern dream interpretation owes much to Sigmund Freud, who theorized that dreams serve two major purposes. First, they preserve sleep by incorporating external stimuli into the dream content. Second, dreams are the expression of the subconscious mind and represent wish fulfillment. Freud emphasized that every dream has a manifest content that is expressed through symbols. To uncover the latent and more important meaning of the dream, one must interpret these symbols, which are largely sexual and universal. Modern dream interpretation builds on, but greatly modifies, Freudian theory. It does so by emphasizing the manifest content of the dream, by supporting a wider range of meaning than wish fulfillment, and by de-emphasizing the universality of symbols. Modern theorists agree with Freud that dreams are the activity of the unconscious mind, but they stress that the unconscious often says what it means, puts forth much useful information, and uses personal symbols that are meaningful only as individuals can interpret them in the context of their own life experiences.

10. The best summary of the present state of research on dreams is that:

F. recent research has proven earlier theorists accurate on most points. A good synthesis of work on dreams is now needed.
G. although relatively recent methods of research have added much knowledge about the dreaming process, there is much to be learned about the physiological as well as the interpretive aspects of dreaming.
H. what is not known about dreams at the present time is probably not worth the research that would be necessary to obtain further information.
J. modern research so completely refutes earlier theories that a breakthrough in dream research seems imminent.

11. What has modern research revealed about sleep?

A. Sleep is cyclic in nature.
B. Dreams take place in an instant.
C. Dreams may be prophetic.
D. Everyone needs at least seven hours of sleep nightly.

12. The author's main intention in the last paragraph is probably to:

F. compare and contrast Freudian dream theory with more modern dream theory.
G. describe the stages of sleep in relation to their dream content.
H. explain the use of symbolism in dream interpretation.
J. define the nature and purpose of dreams.

13. A man has a recurrent dream that his home is on fire. He feels strongly that the meaning of the dream is related to two traumatic events in his childhood. His grandparents' home burned, and his church burned. His analyst insists that fire signifies life, movement, and vitality. They cannot agree about the interpretation of the dream. What is the major difference in their positions?

A. The man is showing the influence of a religious background and is insisting on a prophetic interpretation.
B. The analyst is using a more modern approach to dream interpretation than is the man.
C. The man is using fire as a universal symbol. The analyst sees fire as a personal symbol.
D. The analyst sees fire as a universal symbol. The man is using fire as a personal symbol.

14. A psychologist working in a sleep laboratory designs a study in which she awakens subjects just as they go into the REM stage of sleep. Which area is she most likely studying?

F. Sleep deprivation
G. Dream deprivation
H. Eye movement patterns
J. Dream content

15. Manifest dream content is to latent dream content as literal interpretation is to:

A. direct interpretation.
B. symbolic interpretation.
C. psychic interpretation.
D. exact interpretation.

GO ON TO THE NEXT PAGE.

16. If sleep researchers were to discover that dream deprivation leads to decreased memory, which conclusion about the purpose of dreams would be most logical?

- **F.** Dreams recapitulate events and information collected by the waking mind. This recapitulation may be a part of the memory storage process.
- **G.** Dreams unite the individual with a cosmic consciousness. This union may be one of the bases for memory storage.
- **H.** Dreams provide an outlet for suppressed wishes which, if inhibited, create emotional blocks that decrease memory.
- **J.** Dreams supply the mind with new ways of interpreting remembered information.

17. The history of the study of dreams suggests all of the following predictions EXCEPT that:

- **A.** psychologists will develop increasingly accurate approaches to dream interpretation.
- **B.** studies about the different functions of the two hemispheres of the brain will complement dream study.
- **C.** better understanding of the body's biochemical processes will bring about a clearer understanding of the dreaming process.
- **D.** an intensified emphasis on scientific research will relegate dream study to the area of superstition.

18. Dream deprivation could conceivably be used as a:

- **F.** means to improve mental health.
- **G.** treatment for depression.
- **H.** means to prevent insomnia.
- **J.** brainwashing technique.

GO ON TO THE NEXT PAGE.

3

When the Civil War erupted, many slaves deserted their owners, sought refuge behind federal lines, and appealed to the federal officers for help. In the absence of an official federal policy, each commanding officer formed individual policy according to his needs and position. For example, General Butler, who realized the importance of slave labor to the Confederate war effort, declared runaway slaves "contraband of war" and forbade their return. Ultimately, the officers urged the government to formulate an official policy on this issue.

Such a policy was not established until August 1861, when the Confiscation Act was passed, partly in response to abolitionists' growing dissatisfaction with the federal government's failure to deal decisively with this important issue. This Act stated that any property, including slaves, used with the knowledge or consent of the owner to aid insurrectionists against the United States was lawfully subject to capture and confiscation. The Act specified that confiscated slaves were to be freed, but did not specify what was to be done with them, or for them, thereafter. As the Union army pushed southward and more and more slaves sought refuge behind federal lines, the problem of what to do with the freed slaves increased. In some cases, they were put into special camps, in some cases, they were allowed to farm confiscated lands. In all cases, however, they faced severe problems of readjustment to a completely new way of life. Special commissions and bureaus were formed to find ways and means to help them.

Like the runaway slaves, many of the slaves freed under the Confiscation Act wanted to join the fight, but it was not until the spring of 1862 that Lincoln finally allowed blacks to enlist in the army. Until then, Lincoln had feared the alienation of the border states; he changed his mind at the urging of the officers in the field. During the summer, a few black regiments were formed but were subsequently disbanded. When General Augustus Chetlain assumed control of black recruiting in Tennessee, however, the program became successful. In the following year, serious recruiting efforts began in the Mississippi Valley. Of the total number of black soldiers recruited for the war, more than half were recruited in Confederate states that had fallen to the Union troops. Black regiments became an important link in the chain of Civil War events.

19. As more and more runaway slaves headed toward Northern cities, Northern attitudes toward abolition became increasingly negative primarily because the:

A. neighborhoods that blacks moved into began to deteriorate.
B. blacks could not be assimilated as easily as immigrants.
C. Northerners had wanted to free the slaves but only if they agreed to emigrate.
D. blacks competed with Northern workers for jobs.

20. The Union recognized the right of blacks to be free but did not automatically grant them equal status with whites. This fact is supported by all of the following circumstances EXCEPT:

F. black soldiers being organized into separate, all-black regiments.
G. black soldiers receiving medals of honor for their valorous conduct during the war.
H. black soldiers serving as cooks, stewards, and officers' servants.
J. very few blacks being officers.

21. Which ordering of events is chronologically correct, starting with the earliest event?

I. Emancipation Proclamation
II. Official federal approval of recruitment of blacks
III. Passage of Confiscation Act
IV. Secession of Southern states from the Union

A. II, I, IV, III
B. III, IV, II, I
C. IV, II, III, I
D. IV, III, II, I

22. During the war, Southern slaveholders were probably LEAST likely to use their slaves to:

F. maintain the plantations in the absence of their masters.
G. prepare food for troops.
H. carry messages behind Union lines.
J. help produce ammunition.

23. By 1865 there were many blacks in the North. These free blacks were likely to have been all of the following EXCEPT:

A. slaves who had escaped through the Underground Railway.
B. slaves who had bought their own freedom.
C. slaves who had been freed.
D. foreign immigrants.

24. The Confiscation Act did NOT allow confiscation or capture of a:

F. slave whose master was loyal to the U.S.
G. house in which Confederates had met regularly for strategy sessions.
H. farm on which food for the Confederate army had been grown.
J. slave who had killed his owner's horses rather than let the Union army have them.

25. The Southern states typically assumed a position toward states' rights that was most CONTRADICTED when they:

A. maintained each state's right to determine its own policy on slavery.
B. rejected federal authority over economic policy.
C. upheld the right to secede from the Union.
D. encouraged federal action to recapture fugitive slaves.

GO ON TO THE NEXT PAGE.

26. One problem of readjustment that fugitive slaves did NOT have to face was the:
   - **F.** lack of worldly possessions.
   - **G.** need to learn how to vote.
   - **H.** need to find a job.
   - **J.** need to find a place to live.

27. A runaway slave who sought refuge behind Union lines before 1862 would be LEAST likely to be sent to:
   - **A.** enlist in the army.
   - **B.** farm confiscated territory.
   - **C.** live in a special refugee camp.
   - **D.** find help from special commissions.

28. The Confiscation Act declared runaway slaves contraband if:
   - **F.** their owners had used them in the Confederate war effort.
   - **G.** their owners had declared them fugitives.
   - **H.** they had voluntarily left their owners.
   - **J.** they lived in nonslave states.

Questions 29-43 are not based on a reading passage. You are to answer these questions on the basis of your previous schoolwork in the social studies.

29. The electoral college has been most frequently and seriously criticized because it:
   - **A.** enables a president to be elected with fewer popular votes than his opponent.
   - **B.** is demonstrably inadequate, as proven by the disputed elections of 1800, 1824, and 1876.
   - **C.** is outdated and overly simple.
   - **D.** allows the election of a few weak presidents.

30. Which clause of the Fourteenth Amendment has been the basis for most Supreme Court decisions dealing with the civil rights of minorities?
   - **F.** "All persons born or naturalized in the United States . . . are citizens of the United States."
   - **G.** ". . . when the right to vote is denied to any of the male inhabitants of such State . . . the basis of representation therein shall be reduced."
   - **H.** ". . . nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny . . . equal protection of the laws."
   - **J.** "But neither the United States nor any State . . . shall pay any debt . . . incurred in aid of insurrection . . . against the United States. . . ."

31. All of the following legislative powers check presidential power EXCEPT:
   - **A.** impeachment and removal from office.
   - **B.** changing the size of the Supreme Court.
   - **C.** refusal to confirm presidential appointments.
   - **D.** overriding a veto.

32. An entrepreneur wishes to open a shoe store for women. To determine which community would provide the greatest number of potential customers, she should gather information about:

      I. gender ratios.
     II. total population.
    III. economic statuses.
    IV. competitors.

   - **F.** I and IV only
   - **G.** I, II, and III only
   - **H.** II, III, and IV only
   - **J.** I, II, III, and IV

33. One characteristic of instincts is that they are:
   - **A.** imprinted shortly after birth.
   - **B.** the result of environmental influences after birth.
   - **C.** behavior patterns that vary within a species.
   - **D.** unlearned, inherited behavior patterns.

34. An open primary is one in which voters can:
   - **F.** openly declare allegiance to one party.
   - **G.** participate in the primary of more than one party.
   - **H.** cast their votes without specifying a party preference.
   - **J.** change their parties by pledging to support a new party on election day.

35. One underlying principle of modern cultural studies stresses that every culture must be:
   - **A.** judged in its own context.
   - **B.** visited before it can be properly judged.
   - **C.** judged in terms of economic and industrial progress.
   - **D.** judged in terms of one's own moral standards.

36. The main function of the Federal Reserve System is to:
   - **F.** clear checks for commercial banks.
   - **G.** maintain the gold supply upon which the currency is based.
   - **H.** control the supply of money.
   - **J.** act as a fiscal agent for the federal government.

37. A market-oriented economic system is frequently described as operating subordinate to "consumer sovereignty." This concept means that:
   - **A.** advertisers have little or no impact on sovereign consumers.
   - **B.** producers, rather than consumers, determine the kinds of goods and services produced.
   - **C.** consumers make a careful inventory of their needs for the benefit of producers.
   - **D.** consumer judgments determine the amounts and kinds of goods and services produced.

GO ON TO THE NEXT PAGE.

3

38. The president has great independent authority in foreign affairs. Without the consent of Congress, he can do all of the following EXCEPT:

F. send troops abroad.
G. declare war.
H. proclaim U.S. neutrality.
J. enter into a formal agreement with a foreign nation.

39. The medieval serf's position in society differed from the American slave's position in that the serf:

A. was automatically emancipated upon the death of his master.
B. was exempt from personal, household duties.
C. was forbidden the Christian sacraments.
D. could not be sold away from the land.

40. The New Imperialism of 1870-1914 was characterized by the:

I. search for new industrial markets.
II. need for new sources of raw material.
III. search for new markets for financial investment.
IV. search for strategically located areas to safeguard trade routes.

F. I and II only
G. II and IV only
H. I, II, and III only
J. I, II, III, and IV

41. Jacksonian Democrats were generally opposed to:

A. slavery.
B. the Spoils System.
C. free trade.
D. a centralized national bank.

42. Between World War I and World War II, conditions in Germany and the Soviet Union gave rise to governments that can best be described as:

F. communist.
G. socialist.
H. democratic.
J. totalitarian.

43. Most New England Puritans believed that men of wealth:

A. would never go to heaven.
B. should give away their wealth to the less fortunate.
C. should pay a high tax to the state.
D. should accept wealth as a sign of divine favor.

Since 1940, competition in congressional elections has been declining. At least three major factors contribute to this decline. First, incumbency itself guarantees many advantages: The media routinely report on representatives' actions, so incumbents are better known than their challengers. Incumbents can mail several letters a year to their constituents without paying postage. They are entitled to a number of free trips to their district every year, so there is little danger of their being isolated in Washington.

Second, the securing of committee and subcommittee chairmanships is sometimes influenced by seniority. Having more seniority than any challenger might also occasionally help an incumbent to promote legislation favorable to the district, thus providing voters with great incentive to reelect incumbents.

Third, representatives are increasingly performing valuable, noncontroversial services for their constituents. Since the number both of federal programs and of the people affected by them has mushroomed, mistakes inevitably occur. Some people may fail to receive the assistance due them, such as Social Security payments or health benefits. Others are adversely affected by arbitrary bureaucratic decisions. Faced with the complex regulations surrounding federal programs, they often turn to their representatives for help. And because federal agencies depend upon congressional approval for their budgets and new programs, representatives usually can help: when representatives complain to an agency on behalf of constituents, they nearly always receive quick cooperation. For the incumbent who has performed such services for thousands of constituents, reelection becomes easier.

These trends disturb many observers who believe that representatives spend too much time on constituent services and too little time deliberating legislation. Consequently, they feel that the House may occasionally enact poorly designed programs that create more problems for the people than they solve. The representatives' case loads burgeon, and their attention is further diverted from legislation. Finally, these critics deplore the decline in congressional turnover itself. They argue that a representative's views change little after election to office. Thus the only way to get the House to adopt innovative policies is to elect new representatives, something that is becoming harder to do.

44. Federal agencies are very responsive to representatives' complaints because:

F. the agencies depend upon congressional approval for their budgets and programs.
G. a large number of agency officials are former legislators.
H. agency officials are ultimately responsible to the president.
J. legislators have the right to appoint and to fire agency officials.

GO ON TO THE NEXT PAGE.

45. The political advantages that are enjoyed by incumbent legislators include the opportunity to:

    I. receive a large amount of attention from the mass media.
    II. be closely identified with the federal bureaucracy.
    III. mail literature to constituents postage-free.

    A. I only
    B. II only
    C. I and III only
    D. II and III only

46. The passage can best be characterized as being:

    F. a patriotic tribute to Congress's role in aiding the average citizen.
    G. an openly critical attack on the behavior of several powerful legislators.
    H. an objective report on the status of congressional committees and subcommittees.
    J. a review of factors that may influence congressional competition as well as impair legislative performance.

47. The author states that reelection "becomes easier" for the incumbent who performs many constituent services. What is the underlying assumption here?

    A. Most representatives care about little except getting reelected.
    B. People receiving such services reward incumbents with their votes.
    C. Incumbents are usually better known than their opponents.
    D. Few people are satisfied with the performance of government agencies.

48. Why would "mistakes inevitably occur" in the administration of some federal programs?

    F. Average citizens are demanding more and better government services.
    G. A certain amount of human error enters the implementation of any program.
    H. Certain federal agencies, like the Social Security Administration, are more inept than others.
    J. Federal officials deliberately surround programs with complex regulations in order to discourage public scrutiny.

49. According to the passage, representatives increasingly perform "noncontroversial" services for their constituents. Activities that could be considered controversial would probably include:

    I. speaking at high school graduations and other honorary functions.
    II. sponsoring legislation on issues like tax reform and gun control.
    III. securing federal financing of district water treatment plants.

    A. I only
    B. II only
    C. III only
    D. I and II only

50. According to the passage, critics of Congress believe that the views of most representatives on major issues are:

    F. heavily influenced by the mass media.
    G. dictated by the party's platform.
    H. in agreement with the views of the president.
    J. unlikely to change much over the course of their careers.

51. According to the passage, critics of Congress object to the legislative performance of representatives mainly because:

    A. representatives generally enact programs that are far too big and expensive.
    B. many representatives spend so much time on committee hearings that they often neglect to vote on important legislation.
    C. representatives readily give in to powerful special interest groups.
    D. representatives pay more attention to constituent services than to deliberating new laws.

52. Which of the following constitutional provisions was (were) designed to help keep Congress responsive to changes in public opinion?

    I. Congress can override a presidential veto with a two-thirds vote of both houses.
    II. Congressional elections are held every two years.
    III. Bills pertaining to the collection and expenditure of federal revenue originate in the House of Representatives.

    F. I only
    G. II only
    H. III only
    J. I and II only

END OF TEST 3

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 4
NATURAL SCIENCES READING

35 Minutes—52 Questions

DIRECTIONS: Each passage in this test is followed by several questions. After reading a passage, choose the best answer to each question and blacken the corresponding space on your answer sheet. You may refer to the passage as often as necessary.

When placed together in the same aquarium, male Siamese fighting fish, ***Betta splendens***, exhibit aggressive behavior. Although their aggressive behavior rarely results in prolonged physical contact between fish of the same species, it usually involves a ritual. A biologist performed 4 experiments to learn more about the aggressive behavior of the fighting fish.

Experiment 1

Two aquariums, each containing 1 male Siamese fighting fish, were placed adjacent to each other. As soon as the fish recognized each other, they flared their fins and darkened in color. An occasional attempt at physical contact was noted. After 2 minutes, one male swam to the back of the tank and contracted its fins; also, its skin color lightened. Consequently, the other male stopped any further aggressive behavior.

Experiment 2

When a brightly colored, large female fish was substituted for the male fish in one aquarium, the male in the other aquarium exhibited an aggressive response. When the large female fish was replaced by a lightly colored, small female fish, the male did not exhibit any aggressive behavior. Neither female fish exhibited aggressive behavior toward the male fish. Furthermore, when the 2 female fish were placed in the same tank, they did not exhibit aggressive behavior toward each other.

Experiment 3

To determine which visual cues elicit aggressive behavior in male Siamese fighting fish, painted wooden models were placed in an aquarium containing 1 male Siamese fighting fish. Those models that closely resembled an aggressive male Siamese fighting fish elicited very aggressive responses. Smaller models elicited reduced responses. The more darkly colored the model, the more intense the aggressive response. Models that were darkly colored but had contracted fins elicited very little aggressive action; models that were lightly colored but had extended fins also produced little aggressive action.

Experiment 4

This experiment was designed to determine whether male Siamese fighting fish exhibit territoriality. A male Siamese fighting fish was placed in an aquarium for 1 hour to adjust to its new environment. Then, another male Siamese fighting fish was placed in the same aquarium. This second male was attacked immediately and responded by contracting its fins and lightening in color. This experiment was repeated 5 times with different fish and new water, with the same results. The biologist concluded that male Siamese fighting fish exhibit territoriality.

1. If several male and female Siamese fighting fish were placed together in the same aquarium, one would most likely observe:

A. the darkest and largest male dominating the other fish.
B. breeding behavior.
C. the largest fish attacking the smallest fish.
D. no aggressive behavior among the fish.

2. By turning a light color and contracting its fins when it encounters another male fighting fish, a male Siamese fighting fish is able to:

F. camouflage itself.
G. signal submissiveness and avoid physical harm.
H. look like an immature fish.
J. indicate its readiness to fight.

3. Which experiment supports the hypothesis that a brightly colored female fighting fish exhibits territoriality?

A. Experiment 2
B. Experiment 3
C. Experiment 4
D. None of the above

GO ON TO THE NEXT PAGE.

4. Aggressive behavior is NOT being exhibited when a:
   - F. stickleback fish defends its territory.
   - G. porcupine releases its quills.
   - H. cat arches its back and hisses.
   - J. chameleon turns a dark color.

5. To investigate further the territoriality of male Siamese fighting fish, the experimenter should vary which of the following factors in Experiment 4?
   - A. Temperature of the water
   - B. Size and shape of the aquarium
   - C. Time of day of experimentation
   - D. Salinity of the water

6. Suppose a large female Siamese fighting fish has a permanently dark scale pigmentation. If this mutant female were placed in an aquarium with a male Siamese fighting fish, the male fish probably would:
   - F. ignore the female fish.
   - G. continuously attack the female fish.
   - H. attack the female fish until the female attacked the male.
   - J. attack the female fish until the female flared its fins.

7. Fish species *B* exhibits the same type of aggressive behavior as Siamese fighting fish. One could expect to observe aggressive encounters between:
   - A. males of species *B*.
   - B. females of species *B*.
   - C. a male Siamese fighting fish and a male of species *B*.
   - D. males of species *B* and other types of fish.

8. The experimental data would support the hypothesis that aggressive behavior:
   - F. is not part of the courtship behavior of organisms.
   - G. involves recognizable visual displays.
   - H. is behavior that is innate and not learned.
   - J. is often exhibited between members of 2 different species.

9. Suppose 3 aquariums, each containing 1 male Siamese fighting fish, were placed next to each other so that the fish in each tank could see the fish in every other tank. Which fish would exhibit the most aggressive behavior toward each other?
   - A. The 2 lightest-colored males
   - B. The 2 darkest-colored males
   - C. The darkest- and lightest-colored males
   - D. The 3 fish would not show any aggressive behavior toward each other

10. To clarify the results of Experiment 3, the biologist should use models that:
    - F. show varying degrees of fin contraction and extension.
    - G. have varying girths.
    - H. are made from different materials.
    - J. represent other species of fish.

GO ON TO THE NEXT PAGE.

4

Temperature is a relative measure of hotness or coldness. Temperature changes result from the addition of heat to or the removal of heat from a substance. In the early nineteenth century, the prevailing theory of heat maintained that heat was a fluid, called *caloric*, composed of particles that could flow into and out of an object. The caloric particles were mutually repellent but were attracted with different strengths to all matter. It was also assumed that the fluid was conserved in all interactions. In the caloric theory, the temperature of an object was directly related to the amount of caloric surrounding an atom or a molecule of the object.

In the modern theory of heat, temperature is explained by the concept of molecular kinetic energy. The kinetic theory imagines molecules to be in constant random motion, colliding with each other and the walls of their container. In gases, the distance between molecules is believed to be very large in relation to the size of the molecules. In ideal gases, there are no intermolecular forces to affect the motion of a molecule between collisions. Newton's laws are used to calculate the force exerted by the walls of the container on the gas when the molecules change momentum during the collisions. This force is expressed as pressure (force per area); pressure is related to the volume of the container and the average kinetic energy of the molecules.

The kinetic theory explains the relationships that are observed in nature among the pressure, volume, and temperature of a gas. Temperature (expressed in degrees Kelvin) is directly proportional to the average molecular kinetic energy; the product of pressure and volume is directly proportional to temperature. The kinetic theory also explains a diverse assortment of thermal phenomena. Evaporation is recognized as a cooling process because only the faster moving molecules can break away from the liquid at temperatures below the boiling point, leaving behind the molecules with less kinetic energy and thus lowering the temperature. Most substances expand when heated because the increased kinetic energy of the molecules produces an increase in the average separation distance between molecules.

11. Why is evaporation a cooling process?

- **A.** Molecules of a vapor have less kinetic energy than those of a liquid.
- **B.** Water vapor conducts heat.
- **C.** Evaporation makes the surrounding air damp.
- **D.** The molecules with the greatest kinetic energy escape the liquid.

12. A gas is confined in a container that has rigid walls. When its molecules are made to move faster, how is the gas affected?

- **F.** It condenses to a liquid.
- **G.** Its volume increases.
- **H.** Its temperature increases.
- **J.** Its temperature decreases.

13. According to the caloric theory, the amount of caloric present in a substance per molecule would be a measure of:

- **A.** the temperature of the substance.
- **B.** the mass of the substance.
- **C.** the pressure of the substance.
- **D.** none of the above.

14. When air in a cylinder is compressed by a slowly moving piston, what occurs in this system?

- **F.** The gas cools, because the average kinetic energy of the molecules decreases.
- **G.** The number of molecules increases, because the gas heats up.
- **H.** The temperature of the gas increases, because the air molecules lose energy as they collide with each other.
- **J.** The pressure increases, because the air molecules collide more frequently with the walls of the cylinder.

15. An effective way to increase the number of collisions between the molecules of a gas and the walls of its enclosing container is to:

- **A.** increase the mass of the molecules.
- **B.** increase the velocity of the molecules.
- **C.** increase the size of the molecules.
- **D.** decrease the kinetic energy of the molecules.

16. The force exerted on the walls of a container by a confined gas is directly related to:

- **F.** the changes in the molecules' momentum due to their collisions with the walls.
- **G.** a slight loss in average speed suffered by the molecules at the walls.
- **H.** elastic collisions between the molecules of the gas.
- **J.** the attraction between the walls and the gas.

17. Scientists in the early nineteenth century most probably assumed that caloric repelled itself because they observed that:

- **A.** heat flows from hot objects to cold ones.
- **B.** caloric is conserved.
- **C.** hot objects contain more caloric than cold objects.
- **D.** gases are highly compressible.

18. Which observation suggests that heat is a form of energy, not a substance or fluid?

- **F.** Stirring can warm a liquid.
- **G.** Metal rods conduct heat.
- **H.** Most substances expand when heated.
- **J.** Heat flows from a hot body to a cold one.

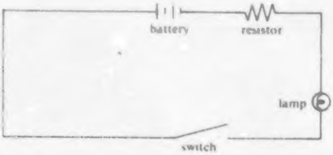
19. If caloric particles existed and had weight, the caloric theory would suggest that as:

- **A.** a quantity of water is heated, its weight decreases.
- **B.** a quantity of lead is cooled, its weight remains the same.
- **C.** an object is heated, its weight increases.
- **D.** a substance is heated, it contracts.

GO ON TO THE NEXT PAGE.

Questions 20-34 are not based on a reading passage. You are to answer these questions on the basis of your previous schoolwork in the natural sciences.

20. In the circuit shown below, what is necessary for the current to light the lamp?



- F. The switch must be closed.
- G. A galvanometer must be attached to the circuit.
- H. Current must travel up and down.
- J. Another cell must be added to the battery.

21. A system of gases is in chemical equilibrium. This equilibrium may be disturbed by an increase in:

- A. pressure
- B. temperature
- C. the amount of reactant
- D. all of the above

22. Mercury is used in many thermometers to measure temperature because mercury is:

- F. the only metal that has a liquid phase
- G. the international scientific standard for measuring temperatures
- H. the only commonly available substance that undergoes measurable changes with temperature
- J. a liquid that increases uniformly in volume as its temperature rises

23. Minerals can be identified by their hardness on a scale from 1 (softest) to 10 (hardest). Hardness tests were made on several minerals by scratching them with a fingernail (hardness 2.5) and a common nail (hardness 5.5). Which mineral could be scratched by the common nail but not by the fingernail?

- A. Gypsum (hardness 2)
- B. Calcite (hardness 3)
- C. Quartz (hardness 7)
- D. Sapphire (hardness 9)

24. On what basis are animals classified into phyla?

- F. Source of energy for locomotion
- G. Type of habitat
- H. Size of torso and limbs
- J. Structure of body parts

25. A student wishes to test the hypothesis that the rate of respiration in humans depends on the amount of available oxygen. He designs a face mask that allows him to rebreathe the same air supply continually; he assumes that the amount of oxygen contained in the air will progressively decrease. The chief error in this procedure is that:

- A. a self-constructed face mask allows too much air to leak out.
- B. the effect of increased carbon dioxide concentration is not accounted for.
- C. the effect of the presence of nitrogen in the air supply is not considered.
- D. the amount of oxygen is not adjusted to account for varying respiration rates in humans.

26. A student conducts an experiment involving a radioactive material and a radiation counter that detects and counts particles of radiation. The student figures that exactly 453 counts will be obtained in 1 minute. This prediction is NOT appropriate because:

- F. it is impossible to calculate the rate of decay of radioactive material.
- G. no radioactive material could produce so many particles in 1 minute.
- H. no radioactive material could produce so few particles in 1 minute.
- J. such a prediction can only be approximate.

27. Fossils are rarely found in igneous rock formations because igneous rocks are:

- A. extremely hard.
- B. scarce.
- C. involved in frequent crustal shifts.
- D. formed at very high temperatures.

28. If only some of the offspring of a pair of fruit flies are vestigial-winged and ebony-colored, what are the genotypes of the parents that were crossed? (v = recessive vestigial-winged; e = recessive ebony-colored; V and E = dominant wild-type characters)

- F. VVEE x vvee
- G. VVEE x VvEe
- H. vvee x vvee
- J. VvEe x VvEe

29. An experimenter adds zinc to hydrochloric acid, and a gas is produced. She collects the gas in a test tube immersed in water and then removes the tube from the water. If she places a glowing splint in the tube, the splint will:

- A. burn vigorously.
- B. glow very brightly.
- C. give off an orange smoke.
- D. cause a popping noise.

GO ON TO THE NEXT PAGE.

30. What do members of the halogen family (fluorine, chlorine, bromine, iodine, and astatine) have in common?

I. Atomic weight
II. Melting point
III. Seven electrons in the outermost electron shell

F. II only
G. III only
H. II and III only
J. I, II, and III

31. Suppose that in an experiment, a bullet is fired from a gun on the surface of the earth, and an identical bullet is fired from a gun on the surface of the moon. Identical guns are used in both situations, and they are fired at identical angles. The bullet that travels further is the one fired on the:

A. moon, because the acceleration due to gravity on the moon is lower than that experienced on the earth.
B. earth, because the bullet travels faster through the earth's atmosphere than through the moon's atmosphere.
C. earth, because the bullet weighs more on the earth than on the moon.
D. earth, because the earth has a greater magnetic field than the moon.

32. To demonstrate osmosis by a rise of fluid within a tube, one should set up a container with a thistle tube and membrane so that:

F. a sugar solution is on both sides of the membrane.
G. a sugar solution is on one side and distilled water is on the other side of the membrane.
H. distilled water is on both sides of the membrane.
J. a sugar solution is on one side of the membrane and the other side of the membrane is empty.

33. A pipette is used in the laboratory to measure:

A. the diameter of small wires.
B. the weight of a solid sample.
C. small distances.
D. a volume of fluid.

34. If red blood cells are placed in concentrated salt solution, they will:

F. mutate.
G. dissolve.
H. fuse.
J. shrink.

Observational studies of star *clusters*, groups of hundreds or thousands of stars bound together by their mutual gravitational attraction, have enabled astronomers to theorize how stars evolve. Since the stars in any cluster all formed simultaneously from the same cloud of interstellar gas and dust, all the stars in a cluster have the same age and chemical composition. However, different clusters have different ages. Thus, by comparing the stars in young clusters to those in old clusters, astronomers can speculate on the processes involved in the life of a star.

Initially, astronomers determine which clusters are young and which are old by considering two simple facts: (1) the amount of fuel a star has available must be related somehow to its mass (large stars start with more fuel than small stars), and (2) a star's luminosity (total rate of radiation) is a direct measurement of its rate of fuel use. Observations show that the largest stars are hundreds of times as massive as, and millions of times more luminous than, the smallest stars. But even though the large stars start with more fuel, they burn out more quickly because they use their fuel much more rapidly than the small stars. Therefore, star clusters that contain a high proportion of bright, massive stars must be young, and those that contain a high proportion of faint, small stars must be old.

Using observational methods to determine the relative ages of clusters, astronomers have found that stars spend most of their lifetimes as *main sequence stars* and that their temperatures and luminosities are directly related to their masses. When the available fuel is exhausted, the star rapidly expands and its surface cools until it becomes a *red giant* or *red supergiant*. After a short time as a giant or supergiant star, the star dies by either fading quietly from view as a *white dwarf* or exploding with tremendous force as a *supernova*. The supernova scatters the material of the star back into space, where it will eventually be recycled into a new generation of stars.

35. In which stage are massive stars hot and bright, and small stars cool and faint?

A. Main sequence
B. Red giant
C. Red supergiant
D. Supernova

36. Which sequence correctly orders the stages that a typical star might undergo during its lifetime?

I. Main sequence
II. Red giant
III. Red supergiant
IV. White dwarf

F. I, II, III
G. I, II, IV
H. II, III, IV
J. II, IV, I

GO ON TO THE NEXT PAGE.

37. On the basis of the information in the passage, which group of stars could be classified as a star cluster?
   - **A.** All stars visible to the naked eye
   - **B.** Two stars in orbit around each other
   - **C.** Stars that pass each other during their wanderings through space
   - **D.** A group of many small and large stars

38. Which statement best describes why star clusters are used to determine how stars evolve?
   - **F.** Stars radiate energy in the form of heat and light.
   - **G.** Star clusters are groups of stars that are gravitationally bound together.
   - **H.** The stars in a cluster are all the same age.
   - **J.** All stars must eventually burn out.

39. If new stars of different masses were formed with equal probability, and new stars were continuously forming and evolving, which statement would be consistent with the passage?
   - **A.** Massive stars should be seen with less probability than small stars, because massive stars do not live as long as small stars.
   - **B.** Massive stars should be seen with more probability than small stars, because massive stars are larger than small stars.
   - **C.** Young stars should be seen with more probability than old stars, because young stars have more energy than old stars.
   - **D.** Young stars should be seen with more probability than old stars, because young stars have a different chemical composition than old stars.

40. Star clusters often contain many more main sequence stars than red supergiants because, as compared to main sequence stars, red supergiants are:
   - **F.** larger.
   - **G.** cooler.
   - **H.** fainter.
   - **J.** more short-lived.

41. Astronomers can compare a star in a single cluster to other stars in the same cluster on the assumption that all stars in a cluster:
   - **A.** die by exploding as supernovas.
   - **B.** have approximately the same chemical composition.
   - **C.** exist for only a few thousand years.
   - **D.** have identical luminosities.

42. Which group of field stars (stars not associated with a cluster) would have the smallest range of ages?
   - **F.** Stars with varying luminosities
   - **G.** All stars within a certain distance from the sun
   - **H.** Very massive stars
   - **J.** Stars millions of years old

43. According to the passage, which single variable determines the length of the main sequence lifetime of a star?
   - **A.** Luminosity
   - **B.** Chemical composition
   - **C.** Mass
   - **D.** Number of stars in the cluster

GO ON TO THE NEXT PAGE.

4

Radioactive decay occurs when nuclei spontaneously change their structures. To determine the characteristics of the *alpha, beta,* and *gamma rays* emitted by these changing nuclei, scientists conducted 3 series of experiments.

First, the capacity of the various emissions to darken photographic plates was observed. Sheets of different materials were placed between the decaying element and the plate. Alpha rays exhibited the least penetrating ability; they failed to penetrate both paper and aluminum sheets. Beta rays, with intermediate penetrating power, passed through several thicknesses of paper and aluminum sheets. But gamma rays penetrated paper, aluminum, and several other materials before they were finally stopped by lead sheets.

Second, the chemical properties of atoms that had undergone radioactive decay were examined. Alpha emissions caused atoms to change to atoms of new elements (called daughter atoms) that had 2 fewer positive charges in the nucleus. For example, uranium 238 emitted 1 alpha particle to become thorium 234, an element whose chemical properties are different from those of uranium. Beta emissions resulted in daughter atoms of elements with either 1 more or 1 less positive charge in the nucleus than the original atoms. Gamma emissions left the chemical properties of atoms unchanged. These experiments also demonstrated that many naturally occurring radioactive atoms undergo a series of radioactive decays, until a stable nucleus (one that does not undergo radioactive decay) is finally produced.

The last series of experiments measured the masses of alpha and beta particles only. Emitted particles (whose velocities had been determined) were passed into a strong magnetic field. The interaction of the moving charge and the magnetic field deflected the particles into circular orbits. The radii of these orbits were proportional to the masses of the particles. Alpha particles, which have the same mass as helium nuclei, were found to have about 7,000 times as much mass as beta particles; beta particles have the same mass as electrons. When daughter nuclei underwent similar processes, those that had decayed by alpha emission had lost 7,000 times as much mass as those that had decayed by beta emission. Those that had decayed by gamma emission had lost no perceptible mass.

44. According to the results of the experiments, which particles would most likely be stopped by a 1-foot layer of water?

I. Alpha particles
II. Beta particles
III. Gamma particles

F. II only
G. III only
H. I and II only
J. II and III only

45. Which of the following charges can a beta particle have?

I. One unit of positive charge
II. One unit of negative charge
III. Two units of positive charge
IV. Two units of negative charge

A. III only
B. IV only
C. I or II only
D. I or IV only

46. Atoms with stable nuclei are found more frequently in nature than atoms of the same element with unstable nuclei. This phenomenon occurs because atoms with:

F. unstable nuclei undergo radioactive decay.
G. unstable nuclei immediately disintegrate into nothing.
H. unstable nuclei readily absorb alpha particles into their nuclei.
J. stable nuclei are usually mixed with radioactive atoms in naturally occurring samples.

47. What do these experiments suggest about the mass of the gamma particle?

A. It equals the mass of an electron.
B. It is so small that it is undetectable.
C. It equals the mass of a helium nucleus.
D. It is greater than the mass of a neutron.

48. On the basis of the first experiment only, one can correctly conclude that radioactive decay:

F. originates in the nuclei of atoms.
G. changes the chemical structure of atoms.
H. occurs in a series.
J. produces a variety of emissions.

49. The experimental results would support which statement about the interaction between moving particles and matter?

A. Uncharged particles are stopped more easily than charged particles.
B. Particles with more mass and more charge are stopped more easily than those with less mass or less charge.
C. Particles with less mass are stopped more easily than those with more mass.
D. Particles with a negative charge are stopped more easily than those with a positive charge.

50. To handle safely a weak, solid radioactive object that emits only alpha particles, what is the minimum precaution that one should take?

F. Wearing plastic gloves and holding the object away from exposed skin
G. Transporting the object in a concrete container
H. Wearing lead-lined gloves and clothing
J. Using mechanical hands to hold the object

GO ON TO THE NEXT PAGE.

51. The third series of experiments depends directly upon which principle?
    - A. Magnetic fields exert forces on moving charges.
    - B. For every action there is an equal and opposite reaction.
    - C. The earth's gravitational force is stronger on more massive particles than on less massive particles.
    - D. The charge of a particle depends on its mass.

52. By emission of which type of particle does the following radioactive decay of carbon occur?

$$^{14}_{6}C \rightarrow {}^{14}_{7}N + {}^{0}_{-1}e$$

    - F. Negative alpha particle
    - G. Positive alpha particle
    - H. Negative beta particle
    - J. Gamma particle

END OF TEST 4

STOP! DO NOT RETURN TO ANY OTHER TEST.

Answer Key to Sample Test Booklet

Test 1 - English Usage

1 B	11 D	21 B	31 C	41 B	51 D	61 D	71 C
2 J	12 G	22 H	32 J	42 G	52 G	62 H	72 H
3 A	13 A	23 D	33 D	43 D	53 D	63 C	73 A
4 H	14 G	24 J	34 H	44 G	54 F	64 G	74 H
5 A	15 B	25 D	35 C	45 B	55 D	65 A	75 C
6 J	16 H	26 H	36 J	46 F	56 H	66 J	
7 B	17 B	27 A	37 A	47 D	57 D	67 B	
8 F	18 J	28 H	38 J	48 F	58 J	68 F	
9 D	19 A	29 D	39 A	49 B	59 A	69 D	
10 H	20 J	30 F	40 H	50 G	60 J	70 H	

Test 2 - Mathematics Usage

1 C	11 B	21 B	31 D
2 K	12 H	22 H	32 H
3 B	13 B	23 D	33 B
4 H	14 G	24 G	34 J
5 E	15 E	25 B	35 A
6 J	16 K	26 G	36 H
7 B	17 B	27 D	37 D
8 F	18 K	28 F	38 G
9 D	19 C	29 C	39 D
10 F	20 F	30 F	40 J

Test 3 - Social Studies Reading

1 D	11 A	21 D	31 B	41 D	51 D
2 J	12 F	22 H	32 J	42 J	52 G
3 A	13 D	23 D	33 D	43 D	
4 F	14 G	24 F	34 H	44 F	
5 A	15 B	25 D	35 A	45 C	
6 G	16 F	26 G	36 H	46 J	
7 D	17 D	27 A	37 D	47 B	
8 H	18 J	28 F	38 G	48 G	
9 B	19 D	29 A	39 D	49 B	
10 G	20 G	30 H	40 J	50 J	

Test 4 - Natural Sciences Reading

1 A	11 D	21 D	31 A	41 B	51 A
2 G	12 H	22 J	32 G	42 H	52 H
3 D	13 A	23 B	33 D	43 C	
4 J	14 J	24 J	34 J	44 H	
5 B	15 B	25 B	35 A	45 C	
6 G	16 F	26 J	36 G	46 F	
7 A	17 A	27 D	37 D	47 B	
8 G	18 F	28 J	38 H	48 J	
9 B	19 C	29 D	39 A	49 B	
10 F	20 F	30 G	40 J	50 F	

Actual Test Directions

This booklet contains four tests—*English Usage, Mathematics Usage, Social Studies Reading*, and *Natural Sciences Reading*. The four tests measure skills and abilities that are highly related to success in college.

During the first half of the testing session, you will complete the English Usage Test and the Mathematics Usage Test. In the second half, you will complete the Social Studies Reading Test and the Natural Sciences Reading Test.

The questions in each test are numbered, and the four or five suggested answers for each question are lettered. On the answer sheet, the rows of ovals are numbered to match the questions and the ovals in each row are lettered to correspond to the suggested answers.

For each question, first decide which answer is best. Next, locate on the answer sheet the row of ovals numbered the same as the question. Then, locate the oval in that row lettered the same as your answer. Finally, blacken the oval completely. Use a soft lead pencil and make your marks heavy and black. *DO NOT USE A BALL-POINT PEN.*

You may work on each test ONLY when your test administrator tells you to do so. If you finish a test before time is called, you should use the time remaining to reconsider questions you are uncertain about in that test. You may NOT look back to a test on which time has already been called, and you may NOT go ahead to another test. To do so will disqualify you from the examination.

Your score on each test will be based only on the number of questions you answer correctly. You will NOT be penalized for guessing. *HENCE IT IS TO YOUR ADVANTAGE TO ANSWER EVERY QUESTION.*

If you change your mind about an answer, erase your first mark thoroughly before marking your new answer. For each question, make certain that you mark in the row of ovals with the same number as the question.

Do not fold or tear the pages of your test booklet. Keep the booklet in as good a condition as possible.

DO NOT OPEN THIS BOOKLET UNTIL TOLD TO DO SO.

The American College Testing Program

The ACT Assessment

Sample Test Booklet*

©1981 by The American College Testing Program. All rights reserved
*Answer key for this sample test booklet is provided on page 35

8121B

TEST 1

ENGLISH USAGE

40 Minutes—75 Questions

DIRECTIONS: In the passages that follow, certain words and phrases are underlined and numbered. In the right-hand column you will find alternatives for each underlined part. You are to choose the one that best expresses the idea, makes the statement appropriate for standard written English, or is worded most consistently with the style and tone of the passage as a whole. Choose the alternative you consider best and blacken the corresponding space on your answer sheet. If you think the original version is best, choose "NO CHANGE." Read each passage through once before you begin to answer the questions which accompany it. You cannot determine most answers without reading several sentences beyond the phrase in question. Be sure that you have read far enough ahead each time you choose an alternative.

Passage I

When I discovered a few months ago, a small cactus struggling [1] to survive in the shade of my patio, I faced a problem that I was unprepared to solve. This cactus was a prickly pear, a variety that thrives in the harsh climatic conditions [2] of deserts. Inside my patio, however, it has been stunted and grows weakly; [3] its pads, or branches, were weak, thin, and yellow. Moreover, included in the evidence was the fact that [4] one pad was damaged by a deep split with a line of gray fiber sealing its edges, [5] and the whole plant leaned forward. Obviously, too much shade and water had weakened it and would break [6] its damaged pad. To save the prickly pear, I decided that moving it was essential; but knowing nothing about transplanting cacti, I called a neighbor to help. He

1. A. NO CHANGE
   B. When I discovered a small cactus struggling a few months ago,
   C. A few months ago, when I discovered a small cactus struggling
   D. When a few months ago, I discovered a small cactus struggling

2. F. NO CHANGE
   G. climatic conditions and temperatures
   H. climatical conditions
   J. climate-induced conditions

3. A. NO CHANGE
   B. was stunted and sickly;
   C. is stunted and grew sick;
   D. grows stunted and weakly;

4. F. NO CHANGE
   G. the evidence, including the fact that
   H. the evidence revealed the fact that
   J. OMIT

5. A. NO CHANGE
   B. it's edges;
   C. its' edge.
   D. their edge;

6. F. NO CHANGE
   G. might have broke
   H. broken
   J. must have broke

GO ON TO THE NEXT PAGE.

broke the prickly pear into two pieces,
uprooted it, laid each in shallow soil
beyond the wall, and covered the roots,
7

adding very little water.
In the weeks following afterwards,
8
the recovery process fascinated me.
The pads drained dry to nourish expanding
roots. When the roots functioned, the
pads refilled. Finally, seeing them erect
9
and green, I knew both transplants had survived.

7. A. NO CHANGE
   B. laid the prickly pear in shallow soil beyond the wall, broke it into two pieces, uprooted it, and covered the roots.
   C. uprooted the prickly pear, broke it into two pieces, laid each in shallow soil beyond the wall, and covered the roots.
   D. broke it into two pieces, uprooted the prickly pear, covered the roots, and laid each in shallow soil beyond the wall.

8. F. NO CHANGE
   G. (Do NOT begin new paragraph) As succeeding weeks followed one another,
   H. (Begin new paragraph) In the succeeding weeks that followed,
   J. (Begin new paragraph) In succeeding weeks,

9. A. NO CHANGE
   B. Finally seeing them standing, erect,
   C. Finally, with them standing, erect
   D. Finally, standing erect

Passage II

America the world's richer agricultural
10
country, was once largely covered with
forests. Needing both timber but also
11
farmland, settlers felled trees and removed
stumps, many of which were ten feet or
more in diameter. They had to remove
these stumps in order to clear the land
for building roads or railroads,
for establishing farming or homes.
12

Imagine standing before a stump
it is big enough to hold a king-size bed.
13

10. F. NO CHANGE
    G. America, the world's richest
    H. America the world's richest
    J. America, the world's most richest

11. A. NO CHANGE
    B. as well as
    C. and
    D. along with

12. F. NO CHANGE
    G. establishing farms or
    H. the establishment of farms or to build
    J. establishing farms, or

13. A. NO CHANGE
    B. being
    C. that has a capacity
    D. OMIT

GO ON TO THE NEXT PAGE.

1

You have a shovel;[14] axes and saws of different sizes. What do you do first? You begin by shoveling a ton or more of dirt away from the visible roots. Then you cut them off below plow level and close to the stump. Since the large taproot goes straight down from the center, one[15] must get down under the stump to cut it. Shoveling, chopping, sawing, and carrying away the pieces that is[16] the process. Finally, when the taproot is almost severed, a good team of horses was[17] able to break the stump loose and roll it out of its hole.[18] Now the stump can be burned. The hole left by the stump[19] can be shoveled full.[20] Only then is that area, a little larger than a king-size bed,[21] ready to be plowed, planted, and harvested. This is what many of the early settlers, stump-ranchers, not sodbusters,[22] had to do to make America a productive agricultural country.

14. F. NO CHANGE
G. shovel
H. shovel and
J. shovel—

15. A. NO CHANGE
B. we
C. you or I
D. you

16. F. NO CHANGE
G. pieces, are
H. pieces—that is
J. pieces being

17. A. NO CHANGE
B. is
C. were
D. OMIT

18. F. NO CHANGE
G. it's hole.
H. its' hole.
J. there.

19. A. NO CHANGE
B. (Do NOT begin new paragraph) The hole of the stump
C. (Begin new paragraph) The stump's hole
D. (Begin new paragraph) The hole, from the stump,

20. F. NO CHANGE
G. and filled in full.
H. full to the brim.
J. in and filled full.

21. A. NO CHANGE
B. bed
C. bed—
D. bed;

22. F. NO CHANGE
G. settlers—stump-ranchers, not sodbusters
H. settlers (stump-ranchers not sodbusters)
J. settlers (stump-ranchers, not sodbusters)

GO ON TO THE NEXT PAGE.

Passage III

The average vacationer seeks relaxation by escaping from daily routine, however,[23] sensitive travelers want more than a change. Hence, they must begin their journeys with:[24] stamina, alertness, and sufficient time to absorb the information they gather.

Any traveler must learn about local transportation and facilities, but the traveler who seeks knowledge must also learn national laws and such local customs like which[25] proper side of the road to drive on or the proper amount to tip the maid, although[26] it is only human to observe other people, such observation should be tactful. People want[27] above all, to satisfy their curiosity about their own species. Therefore, to begin a journey with preconceived ideas of what is good or bad, beautiful or ugly, worthwhile and not useless[28] demeans not only the traveler, but also[29] the unfamiliar culture.

The ability to absorb impressions comes[30] with experience and prevents the traveler from being neither overawed by or[31] bored with new scenes and customs. Sensitive travelers never tire of their journeys.

23. **A.** NO CHANGE
B. routine however,
C. routine.
D. routine, but

24. **F.** NO CHANGE
G. with such traits as the following
H. with such traits as
J. with

25. **A.** NO CHANGE
B. like
C. as the
D. as to which

26. **F.** NO CHANGE
G. maid, because
H. maid although
J. maid. Although

27. **A.** NO CHANGE
B. People want,
C. People wanting
D. People, wanting

28. **F.** NO CHANGE
G. what is worthwhile or useless
H. worthwhile or useless
J. worthwhile and useful

29. **A.** NO CHANGE
B. traveler and
C. traveler but in addition
D. traveler, and also

30. **F.** NO CHANGE
G. impressions, which come
H. impressions come
J. impressions coming

31. **A.** NO CHANGE
B. either overawed by or
C. neither overawed by and
D. either overawed by and

GO ON TO THE NEXT PAGE.

1

They are always refreshed and fulfilled by it and [32] of course, are always eager to travel again, for they seek a kinship with all people.

32. F. NO CHANGE
    G. them,
    H. them and,
    J. it,

Passage IV

The recent emphasis on the equality and dignity of all persons drawing [33] attention to job titles. In many industries, the names of jobs have been changed, although [34] the functions performed by the jobholders have not. How long, for example, has it been since there has been an advertisement for janitors? Today he has been [35] referred to as custodial or sanitary engineers.

The bureaucratic reasoning behind these changes suggests more than mere window dressing, its [36] not just that "longer is always better." Some of the changes are meant to upgrade the job and give the persons performing for it [37] an appropriate sense of dignity. To the extent, that [38] new titles can do this, fine; to the extent that they cannot, they should be rejected.

33. A. NO CHANGE
    B. can draw
    C. have drawn
    D. has drawn

34. F. NO CHANGE
    G. changed, and
    H. changed, therefore
    J. changed, however

35. A. NO CHANGE
    B. they have been
    C. he is
    D. they are

36. F. NO CHANGE
    G. dressing it's
    H. dressing. Its
    J. dressing; it's

37. A. NO CHANGE
    B. for them
    C. within them
    D. it

38. F. NO CHANGE
    G. extent that
    H. extent by which
    J. extent where

GO ON TO THE NEXT PAGE.

A connected[39] related shift involves changing the job title to remove sexist overtones. Until recently, for example, many job titles contain[40] a reference to the sex of the person who most frequently performed that job. Since the laws for[41] discrimination in hiring have finally been passed, it is fitting that not only legal but also verbal barriers to jobs be eliminated.

39. A. NO CHANGE
    B. connected,
    C. connected but
    D. OMIT

40. F. NO CHANGE
    G. contained
    H. containing
    J. have

41. A. NO CHANGE
    B. on
    C. about
    D. against

Passage V

Established under the terms of Alfred Nobel's will, the Nobel Prize, of all literary awards is[42] the most prestigious. Having made a fortune from explosives, Nobel regretted his[43] violent consequences and seeking[44] to promote constructive and creative[45] not destructive, efforts.

Assuring its[46] winner fame and financial success, the literary award is highly coveted. The laureate, or winner, receives

42. F. NO CHANGE
    G. awards, is
    H. awards are
    J. awards, are

43. A. NO CHANGE
    B. their
    C. that
    D. one's

44. F. NO CHANGE
    G. having sought
    H. in seeking
    J. sought

45. A. NO CHANGE
    B. creative and
    C. creative but
    D. creative,

46. F. NO CHANGE
    G. it's
    H. its'
    J. their

GO ON TO THE NEXT PAGE.

a medal, a certificate, and approximately $100,000; and, customarily, also delivering [47] a speech.

Though acclaimed, laureates are [48] often criticized because they have not always been popular choices. For example, America's first laureate, Sinclair Lewis, who created an outrage [49] with his novels critical of middle-class values. Pearl Buck's selection, [50] whose Chinese peasant novels had not established her among critics as "literary," also evoked [51] a storm of protest. John Steinbeck was not considered a first-rate writer by the literati. That his choice was peculiar is shown by the fact that it created [52] little public attention, suggesting that he was not widely read in 1962. Finally, Americans viewed the novels of William Faulkner as esoteric, obscene, and unreadable. [53]

Some American recipients have been popular and dramatic choices. The choice of O'Neill was applauded, like Eliot's, [54] whose poetry of fragmented man captured the spirit of the twenties. In 1976, the choice of Chicagoan Saul Bellow created not only a stir because of his urban, rootless heroes, and [55] a "first": all six Nobel awards went to Americans.

47. A. NO CHANGE
    B. $100,000, customarily, he also delivers
    C. $100,000. And, customarily, also delivers
    D. $100,000. Customarily, he also delivers

48. F. NO CHANGE
    G. (Begin new paragraph) Though acclaimed laureates,
    H. (Begin new paragraph) Though acclaiming laureates,
    J. (Do NOT begin new paragraph) Though acclaiming laureates,

49. A. NO CHANGE
    B. created an outrage of people
    C. creating an outrage
    D. outraged people

50. F. NO CHANGE
    G. Pearl Buck's winning,
    H. The Pearl Buck choice,
    J. The selection of Pearl Buck,

51. A. NO CHANGE
    B. projected
    C. was evocative of
    D. was provocative of

52. F. NO CHANGE
    G. to have created
    H. by creating
    J. due to its creating

53. A. NO CHANGE
    B. as obscene, and hard to read.
    C. obscene, as well as being unreadable.
    D. obscene, and judged unreadable.

54. F. NO CHANGE
    G. as was Eliot's poetry,
    H. as was that of Eliot,
    J. like Eliot,

55. A. NO CHANGE
    B. as well as
    C. and also
    D. but also

GO ON TO THE NEXT PAGE.

Passage VI

James Baldwin, a major American writer who [56] can be called a twentieth-century Amos. The biblical prophet Amos lived during an age and time [57] of tremendous cultural advancement and economic prosperity that was accompanied by moral and religious corruption. Baldwin feels to believe [58] that he has seen a similar decline in America. Amos was extremely sensitive, and aware, [59] of the shortcomings of the kingdom that he addressed. Baldwin (also) has [60] demonstrated keen perception in assessing the significance of social, moral, and political trends in America. In fact, in his writings over the years, he has accurately defined America's problems. However, like Amos, [61] Baldwin has protested the actions that he felt indicated a lack of respect and adherence for [62] the laws of God: the luxurious life-styles of the wealthy that were supported by their oppressive economic exploitation of the poor, the lack of [63] responsibility of the privileged for the less privileged, and the need for justice for all people. Critics severely denounced both Amos and Baldwin for their strong stands on these issues.

The predictions of both men were consistently gloomy; so far, Baldwin's

56. F. NO CHANGE
G. A major American writer, James Baldwin,
H. James Baldwin is a major American writer, who
J. James Baldwin, a major American writer,

57. A. NO CHANGE
B. in a time
C. in time
D. OMIT

58. F. NO CHANGE
G. believes
H. in his opinion believes
J. feels and believes

59. A. NO CHANGE
B. (Do NOT begin new paragraph) Amos was extremely sensitive
C. (Begin new paragraph) Amos was painfully aware
D. (Begin new paragraph) Amos was extremely sensitive

60. F. NO CHANGE
G. Baldwin, also has
H. Baldwin also has
J. Baldwin has, also

61. A. NO CHANGE
B. Nevertheless, like Amos,
C. Similarly, like Amos
D. Like Amos,

62. F. NO CHANGE
G. for and adherence to
H. as well as adherence to
J. and a lack of adherence to

63. A. NO CHANGE
B. poor. The lack of
C. poor—the lack of
D. poor, to lack

GO ON TO THE NEXT PAGE.

predictions have been just as consistently accurate. Many of his critics have labeled him the "prophet of doom"; accordingly 64 black Americans, young and old, regarding him highly 65 as a leader in their struggle for change.

64. F. NO CHANGE
G. since
H. furthermore
J. but

65. A. NO CHANGE
B. regard him
C. look up to him
D. have great regard for him

Passage VII

The reason that 66 bicycles are becoming a popular mode of transportation are obvious: they're 67 good for the environment and for your health. But to have fun cycling, you must choose your bike carefully. 68 If it is not the right size, you will become tired (exhausted) 69 and uncomfortable. An improperly fitted bicycle will also increase the danger of your losing control.

However, 70 a bike that is either too heavy or cumbersome will reduce your pleasure and efficiency, so that you have spent 71 much energy to gain little ground.

In addition to questions of 72 size and weight, you must consider that of gearing.

66. F. NO CHANGE
G. The reasons that
H. The reason why
J. Why

67. A. NO CHANGE
B. obvious; their
C. obvious, there
D. obvious; its

68. F. NO CHANGE
G. care must be used in choosing your bike.
H. the bike you choose must be selected with care.
J. the bike must be carefully chosen by you.

69. A. NO CHANGE
B. tired
C. tired and exhausted
D. tired, and exhausted.

70. F. NO CHANGE
G. (Begin new paragraph) Since
H. (Do NOT begin new paragraph) Since
J. (Do NOT begin new paragraph) Furthermore,

71. A. NO CHANGE
B. you spent
C. you are spending
D. you'll spend

72. F. NO CHANGE
G. questions over
H. questioning about
J. questions for

GO ON TO THE NEXT PAGE.

The factors that should influence your decision includes:[73] the local terrain, the amount of cycling you plan to do, and the types of weather you'll face. Although you might find it difficult to weigh all these,[74] in choosing a bicycle, most bike-shop owners are themselves enthusiastic riders and are more than willing to give free advice, even if you don't buy a bike on his account.[75]

73. A. NO CHANGE
    B. decision, includes
    C. decision:
    D. decision include

74. F. NO CHANGE
    G. all these factors,
    H. all these factors
    J. them

75. A. NO CHANGE
    B. due to them.
    C. from them.
    D. particularly from their store.

END OF TEST 1

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 2

MATHEMATICS USAGE

50 Minutes—40 Questions

DIRECTIONS: Solve each problem, choose the correct answer, and then blacken the corresponding space on your answer sheet.

Do not linger over problems that take too much time. Solve as many as you can; then return to the others in the time you have left.

Note: Illustrative figures are not necessarily drawn to scale and are assumed to lie in a plane unless otherwise stated.

DO YOUR FIGURING HERE.

1. A recipe calls for $1\frac{1}{2}$ cups of flour to be used with $6\frac{1}{2}$ cups of milk. If only $\frac{3}{4}$ of a cup of flour is to be used, how many cups of milk are needed?

   A. 2

   B. 3

   C. $3\frac{1}{8}$

   D. $3\frac{1}{4}$

   E. $3\frac{1}{2}$

2. The relationship between the number of times a cricket chirps in a minute (n) and the temperature (t) is expressed by the formula $t = \frac{n}{4} + 40$. If a cricket chirps 56 times in a minute, what is the temperature?

   F. 16°
   G. 24°
   H. 54°
   J. 64°
   K. 96°

3. Which of the following statements is NOT true?

   A. $5(a - 8) = 5a - 40$

   B. $7(6 + \frac{2}{7}) = 42 + 2$

   C. $4(2b + 2) = 8(b + \frac{1}{2})$

   D. $5b - 20 = 5(b - 4)$

   E. $2(\frac{1}{2}m - n) = m - 2n$

4. The average weight of 3 boxes is 18 grams. If one box weighs 30 grams and another weighs 15 grams, what is the weight, in grams, of the third box?

   F. 9
   G. 12
   H. 15
   J. 18
   K. 21

GO ON TO THE NEXT PAGE.

5. $|-6| - |-5| = ?$
   - **A.** −11
   - **B.** −1
   - **C.** 1
   - **D.** 11
   - **E.** None of the above

6. Helen discovers that when she multiplies the number stored in the memory of her calculator by 6 she gets a value that is twice as large as she gets when she adds 12 to the number in the memory. What is the number she has stored in the memory of the calculator?
   - **F.** 1.0909
   - **G.** 2.1818
   - **H.** 3.0000
   - **J.** 4.0000
   - **K.** 6.0000

7. Mr. Green paid $7.00 for a can of paint and a brush. If he had bought 6 cans of paint and 2 brushes, his bill would have been $21.00. How much did the brush cost?
   - **A.** $ 1.75
   - **B.** $ 3.50
   - **C.** $ 4.20
   - **D.** $ 5.25
   - **E.** $10.50

8. 0.0833 ÷ 3.5 = ?
   - **F.** 0.0238
   - **G.** 0.2380
   - **H.** 2.6670
   - **J.** 3.4167
   - **K.** 4.2010

9. If $t = 3$ and $r = 2$, then $\frac{tr^3}{(t-r)^3} = ?$
   - **A.** 216
   - **B.** $\frac{216}{5}$
   - **C.** 24
   - **D.** 18
   - **E.** $\frac{18}{5}$

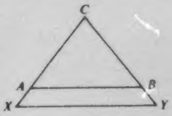
10. For what value of k is the average of 3, 5, and k equal to 6?
    - **F.** −6
    - **G.** −2
    - **H.** $\frac{2}{5}$
    - **J.** $\frac{12}{5}$
    - **K.** 10

DO YOUR FIGURING HERE.

GO ON TO THE NEXT PAGE.

11. In $\triangle CXY$ below, $\overline{AB} \parallel \overline{XY}$, $\angle CAB$ measures 52°, and $\overline{CX}$ and $\overline{CY}$ are of equal length. What is the measure of $\angle ACB$?

DO YOUR FIGURING HERE.



- **A.** 50°
- **B.** 52°
- **C.** 76°
- **D.** 104°
- **E.** 128°

12. $\left[\left(\frac{1}{2}\right)^3\right]^2 = ?$

- **F.** $\frac{1}{32}$
- **G.** $\frac{1}{64}$
- **H.** $\sqrt{\frac{1}{2}}$
- **J.** $\sqrt{\frac{1}{4}}$
- **K.** $\sqrt{8}$

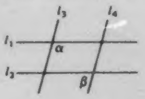
13. What is the largest prime factor of 42 ?

- **A.** 2
- **B.** 3
- **C.** 7
- **D.** 13
- **E.** 42

14. Jack buys a shirt for $9.00, a tie for $5.00, and a pair of slacks. If the 4 percent sales tax on the total purchase amounts to $1.04, how much did his slacks cost?

- **F.** $12.00
- **G.** $13.00
- **H.** $14.00
- **J.** $15.04
- **K.** $26.00

15. In the figure below, if $l_1 \parallel l_2$, $l_3 \parallel l_4$, and the measure of $\angle\alpha = 105°$, then the measure of $\angle\beta = ?$



- **A.** 75°
- **B.** 105°
- **C.** 135°
- **D.** 140°
- **E.** 150°

GO ON TO THE NEXT PAGE.

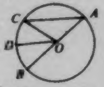
DO YOUR FIGURING HERE.

16. Mr. Peterson earns a base salary plus a 17% commission on his sales. One week he sold $700 of merchandise and earned a total of $250. How much is Mr. Peterson's base salary per week?
    - F. $ 76.50
    - G. $131.00
    - H. $233.00
    - J. $369.00
    - K. $657.50

17. If $f(x) = 3x - 2$, then $\frac{1}{9}[f(x) + 8] = ?$
    - A. x
    - B. $\frac{1}{3}(x - 2)$
    - C. $\frac{1}{3}(x + 2)$
    - D. $\frac{1}{3}(x + 18)$
    - E. $3x - 2$

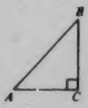
18. A and B are two sets where $A = \{x \mid 1 \leq x < 4\}$ and $B = \{0, 1, 2, 3, 4\}$. What is the intersection $(A \cap B)$ of the two sets?
    - F. $\{1,2\}$
    - G. $\{1,2,3\}$
    - H. $\{1,2,3,4\}$
    - J. $\{x \mid 1 \leq x < 4\}$
    - K. $\{x \mid 1 \leq x \leq 4\}$

19. In the diagram below, O is the center of the circle, and $\overline{AB}$ is a diameter. $\overline{OD}$ bisects $\angle BOC$. Compared to the measure of $\angle BAC$, the measure of $\angle BOD$ is:



    - A. one-third as large.
    - B. half as large.
    - C. equal.
    - D. twice as large.
    - E. three times as large.

20. In right triangle ACB, if $\sin A = \frac{4}{5}$ and the length of $\overline{BC} = 24$, how long is $\overline{AC}$?



    - F. 12
    - G. 16
    - H. 18
    - J. 20
    - K. 25

GO ON TO THE NEXT PAGE.

2

DO YOUR FIGURING HERE.

21. Bill averaged 34 kilometers per hour for the first 2 hours of his motorcycle trip and traveled the last 250 kilometers in 4 hours. What was Bill's average speed, in kilometers per hour, for the entire trip?
    - **A.** 47.33
    - **B.** 48.25
    - **C.** 53.00
    - **D.** 71.00
    - **E.** 96.50

22. A numeration system has a as its base. If $4_a + 4_a + 4_a = 20_a$, then $24_a + 24_a + 24_a = ?$
    - **F.** 72_a
    - **G.** 80_a
    - **H.** 82_a
    - **J.** 120_a
    - **K.** 200_a

23. In $\triangle ABC$, if the length of $\overline{AB} = 5x^3 + 3x^2 + 5x$, the length of $\overline{BC} = 7x^3 + 7x^2 - 8x$, and the length of $\overline{AC} = 3x^3 - x^2$, then for some real number, x, the perimeter of $\triangle ABC = ?$
    - **A.** $x(15x^2 + 11x - 3)$
    - **B.** $x(15x^2 + 9x + 13)$
    - **C.** $x(15x^2 + 11x + 13)$
    - **D.** $3x(5x^2 + 3x - 1)$
    - **E.** $3x(5x^2 + 3x)$

24. A number, x, is decreased by 7; when the result is subtracted from 12, the remainder is 3. Which equation could be used to find x?
    - **F.** $7 - (12 - x) = 3$
    - **G.** $12 - (x - 7) = 3$
    - **H.** $(12 - x) - 7 = 3$
    - **J.** $12 - (7 - x) = 3$
    - **K.** $(x - 7) - 12 = 3$

25. $\dfrac{2^8[(2)(2^6)]}{2^3} = ?$
    - **A.** 2^{17}
    - **B.** 2^{16}
    - **C.** 2^{12}
    - **D.** 2^{11}
    - **E.** 2^5

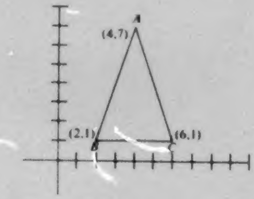
26. Mr. Jansen's take-home pay is 75% of his salary. If his take-home pay is $286.50, what is his salary (to the nearest dollar)?
    - **F.** $190
    - **G.** $215
    - **H.** $358
    - **J.** $382
    - **K.** None of the above

27. $\dfrac{\sqrt{6} - \sqrt{27}}{\sqrt{3}}$ equals which of the following?
    - **A.** -7
    - **B.** $-\sqrt{7}$
    - **C.** $\sqrt{2} - 3$
    - **D.** $\sqrt{7}$
    - **E.** $3\sqrt{6}$

GO ON TO THE NEXT PAGE.

DO YOUR FIGURING HERE.

28. What is the area of ΔABC?

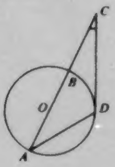


F. 4
G. 6
H. 8
J. 12
K. 24

29. What is the length in centimeters of the hypotenuse of a right triangle if the lengths of its legs are 9 centimeters and 4 centimeters?

A. $\sqrt{13}$
B. 5
C. $\sqrt{97}$
D. 13
E. 97

30. In the figure below, circle O has diameter $\overline{AB}$ and tangent $\overline{CD}$ at D. If the measure of $\angle BAD = 25°$, what is the measure of $\angle BCD$?



F. 25°
G. 40°
H. 65°
J. $77\frac{1}{2}$°
K. 80°

31. The set of all numbers, x, such that $x^2 - 1 < 3$ is equal to the set of all numbers, x, such that:

A. $-\sqrt{2} < x < \sqrt{2}$
B. $-2 < x < 2$
C. $x < -2$ or $x > 2$
D. $x < 2$
E. $x > -2$

GO ON TO THE NEXT PAGE.

32. If $x - y + 8 = 0$ and $13x - 5y = 0$, then $\sqrt{y^2 - x^2} = ?$

F. ±5
G. ±8
H. ±12
J. ±13
K. ±25

DO YOUR FIGURING HERE.

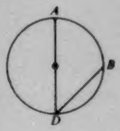
33. $\{x \mid x^2 - 3x = 10\} = ?$

A. {-10,1}
B. {-5,2}
C. {2,5}
D. {5,-2}
E. {10,1}

34. In a particular basketball game Sue scored $\frac{1}{3}$ of the team's points, and Tammy scored another $\frac{1}{4}$ of the team's points. If Sue scored 4 more points than Tammy did, how many points did the team score?

F. 14
G. 24
H. 30
J. 36
K. 48

35. In the circle below with diameter $\overline{AD}$, the length of $\overline{BD} = 2\sqrt{2}$ and the measure of arc $\widehat{AB} = 90°$. What is the radius of the circle?



A. 2
B. $2\sqrt{2}$
C. 4
D. $4\sqrt{2}$
E. 8

36. The product of 2 integers is negative. Their sum is positive. How small could the larger of the two integers be?

F. -2
G. -1
H. 0
J. 1
K. 2

37. If Jane is 4 years younger than Mary, and Mary is $1\frac{1}{2}$ times as old as Jane, what is the product of their ages?

A. 12
B. 45
C. 54
D. 96
E. 165

GO ON TO THE NEXT PAGE.

DO YOUR FIGURING HERE.

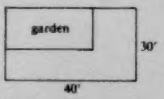
38. When Beth was 1 month old, she weighed [illegible] of her birth weight. At 2 months, she weighed [illegible] of her 1-month weight, and at 3 months she weighed [illegible] of her 2-month weight. If her 3-month weight is 12.5 pounds, how many pounds did she weigh at birth?

- **F.** 3.60
- **G.** 5.00
- **H.** 6.25
- **J.** 6.40
- **K.** 8.00

39. If $\log_{10}2 = .301$ and $\log_{10}3 = .477$, then $\log_{10}72 = ?$

- **A.** 1.556
- **B.** 1.857
- **C.** 2.033
- **D.** 2.334
- **E.** 2.635

40. An ecologist owns the 30′ x 40′ rectangular lot shown below. She wants to plant a rectangular garden that will take up half the area of the lot. If the length of the garden is 13 feet longer than the width, x, what equation could be used to find the width of the garden?



- **F.** $x^2 + 13x = 600$
- **G.** $x^2 + 13 = 600$
- **H.** $2x + 13 = 40$
- **J.** $2(x + 13) = 40$
- **K.** $x(x + 13) = 1,200$

END OF TEST 2

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 3
SOCIAL STUDIES READING

35 Minutes—52 Questions

DIRECTIONS: Each passage in this test is followed by several questions. After reading a passage, choose the best answer to each question and blacken the corresponding space on your answer sheet. You may refer to the passage as often as necessary.

Congress recently passed legislation requiring the government to fund the election campaigns of presidential candidates. The following speakers discuss this legislation.

Speaker 1

Private campaign contributions are the major source of corruption in American politics. Large campaign contributions are traded for favorable legislation and other favors. Rather than debate the issues, candidates offer themselves for sale to the highest bidder. Furthermore, the system of private contributions imposes no ceiling on spending. Candidates can win simply by outspending their rivals. The costs of campaigning have consequently soared. Public campaign financing will eliminate the corruption and reduce the wasteful expense of the present system.

Speaker 2

Free, privately controlled areas of life have increasingly come under government control. Now self-styled reformers, oblivious to this creeping socialism, want to expand government further by having it finance election campaigns. Private campaign financing has worked well. Naturally, interest groups from business to labor, industry to environmentalists—contribute generously to candidates supporting their views. But this insures that candidates representing widely differing viewpoints are well financed and can conduct vigorous campaigns. Public financing, however, discourages interest group participation and undermines the pluralistic, competitive nature of American politics.

Speaker 3

Since public campaign financing would substantially limit the need for private contributions, opponents argue that it limits freedom of speech. Campaign contributions, they say, are indications of support, and thus acts of speech. Perhaps theoretically valid, this argument is wrong in practice. Too often contributions pay for special favors. If contributions are acts of speech, why are so many given secretly? Public campaign financing would remove the unfair advantage held by the rich. Rather than suppress free speech, it will help give everyone an equal voice.

Speaker 4

Public campaign financing has one overwhelming drawback. Whenever government handouts are offered, crowds always gather to get a share of them. If the government pays candidates to campaign for office, the number of candidates—especially one-issue and frivolous candidates—will increase dramatically. It is unfair to taxpayers to institute another government giveaway.

1. By "one-issue and frivolous candidates," Speaker 4 means candidates who:
   - **A.** have no chance of winning but run only to attract publicity for themselves and their causes.
   - **B.** run for unimportant, honorific offices.
   - **C.** lack a broad educational background.
   - **D.** are ineligible to receive government payments under public campaign financing.

2. When Speaker 1 declares that "the system of private contributions imposes no ceiling on spending," she means that:
   - **F.** most candidates finance their campaigns with large contributions from a few private sources.
   - **G.** modern campaigns are more costly than campaigns in the past.
   - **H.** candidates can spend as much as they wish on their campaigns.
   - **J.** almost all candidates need vast personal funds in order to run for office.

3. When Speaker 1 claims that "candidates offer themselves for sale to the highest bidder," she means that political candidates:
   - **A.** simply pay voters to vote for them.
   - **B.** spend enormous amounts of money on their election campaigns.
   - **C.** are more interested in making money than in winning elections.
   - **D.** tailor their positions on issues to attract large private campaign contributions.

4. Which of the following speakers would be most likely to favor legislation requiring candidates to publicly disclose the sources of all campaign contributions?
   - **F.** 1
   - **G.** 2
   - **H.** 1 and 3
   - **J.** 2 and 4

GO ON TO THE NEXT PAGE.

5. What is the "unfair advantage" of the rich to which Speaker 3 refers?
   - A. They would receive special favors under public campaign financing.
   - B. They can better afford to run for public office.
   - C. Public campaign financing would limit private contributions and thus benefit the rich disproportionately.
   - D. They can make much larger campaign contributions and thus have more political influence.

6. When Speaker 2 states that advocates of public campaign financing are "oblivious to this creeping socialism," he means that they:
   - F. are unduly influenced by special interest groups.
   - G. are being used by socialists to further socialist aims.
   - H. fail to see the danger of too much government control.
   - J. seriously underestimate the high financial costs of campaigning.

7. Which speaker is most concerned about the expense of public campaign financing to the federal government?
   - A. 1
   - B. 2
   - C. 3
   - D. 4

8. If Speaker 4's argument against public campaign financing is valid, which rule might lawmakers institute to meet his objections?
   - F. Candidates will not receive subsidies unless they collect a large number of signatures from supporters in all fifty states.
   - G. Candidates can supplement their public subsidies with private contributions if they wish.
   - H. Candidates will have to forfeit a dollar in public subsidies for every dollar of private contributions they receive.
   - J. Candidates must make public the names of all individuals and organizations contributing to their campaigns.

9. By voicing opposition to government "handouts" and "giveaways," Speaker 4 implies that he is opposed to:
   - A. allowing government employees to contribute money to political candidates.
   - B. high levels of government spending in general.
   - C. the "wasteful expense of the present system" to which Speaker 1 refers.
   - D. the practice of holding frequent elections.

GO ON TO THE NEXT PAGE.

3

America's cities have traditionally been the springboard that launched immigrant groups into the mainstream of our culture. Crowded into miserable slums, new arrivals took menial jobs and struggled to educate their children. The second or third generation achieved upward mobility through economic advancement and political organization. Most were able to merge eventually into the general population and leave the slums. The Irish, Germans, Jews, Italians, and East Europeans each in turn followed this process until legislation restricted foreign immigration in the 1920s.

But the flow of new people to repopulate urban slums continued unabated. Harassed and disheartened by disenfranchisement and the imposition of Jim Crow laws in the South at the turn of the century, blacks began to move to northern cities in large numbers. Employment opportunities generated there by World Wars I and II attracted them, as did the chance for their children to get a better education.

As had been the case with foreign immigrants, blacks moved into the slums, took the unskilled, low-paying jobs, and placed their faith in education. But things had changed. Industrialization had reduced the need for unskilled labor. Public policy decisions favoring government-supported home loans and subsidized highway systems encouraged the affluent to move to the suburbs. With the urban tax base eroding, schools and urban conditions in general deteriorated. Furthermore, racial discrimination against blacks was particularly harmful in housing. Blacks who saved their money and were able to afford better homes found none available to them. They were restricted to the ghetto.

Fortunately, the ballot box in the North was colorblind. White flight combined with black migration made the black urban vote a significant factor in national as well as local elections. Organization into a bloc vote had been a standard technique used by foreign immigrant groups, and blacks adopted the same tactic.

By the 1960s, several major American cities had elected black mayors. Reapportionment of congressional districts resulted in the election of blacks to the House of Representatives. Black legislators served on important committees and were able to influence the federal bureaucracy to respond to their needs. The city thus began to serve Afro-Americans in the same fashion as it had earlier served foreign immigrants.

10. Which factor best accounts for the success of black political efforts in recent years?

- **F.** Republicans have successfully recruited many black candidates.
- **G.** Democrats have generally held the governorships in a majority of the states.
- **H.** The Supreme Court has required that a certain percentage of elected officials must be black.
- **J.** The number of black voters in urban areas has increased enough for blacks to control local politics to some degree.

11. Which of the following government policies encouraged people to move from cities to suburbs?

- **A.** The federal highway programs, which enabled people to live at some distance from the city and still commute to work
- **B.** Legislation that made interest payments on home mortgages deductible from income tax
- **C.** Government guarantees for mortgages, which brought home ownership within the reach of the lower middle class
- **D.** All of the above

12. Which situation best describes one effect of housing discrimination in urban areas?

- **F.** Because only inferior, but low-rent, housing was available to blacks, they were able to devote more of their income to other needs.
- **G.** As more blacks competed for the limited housing units available to them, rents increased and the percentage of income necessary to provide shelter rose markedly.
- **H.** As rents rose for inner-city apartments, housing became a profitable investment, and private industry responded by undertaking major construction projects in black urban neighborhoods.
- **J.** As neighborhoods became identified with one race or the other, pressures were relaxed and the rate of violent crime decreased.

13. At the end of Reconstruction in 1877, one of the major means of disenfranchising southern blacks was the

- **A.** grandfather clause, which permitted only older blacks to vote.
- **B.** literacy test, which was subjective and administered by white officials.
- **C.** requirement that age be documented, which few blacks could do.
- **D.** residence requirement, which few blacks could meet.

14. How does the flight to the suburbs affect the urban tax base?

- **F.** It eliminates from the urban population those people best able to finance necessary city services.
- **G.** The rise of suburban shopping centers reduces congestion downtown.
- **H.** Those who remain in the city need very few city services.
- **J.** Those who move to the suburbs are mostly middle-aged, affluent people whose children are grown, and who thus need few city services.

15. At the turn of the century, which difficulties did blacks have in common with eastern European immigrants?

I. They had to master a new language.
II. Their agrarian background did not give them urban job skills.
III. They lacked formal education.
IV. They faced religious prejudice.

- **A.** I and III only
- **B.** II and III only
- **C.** III and IV only
- **D.** II, III, and IV only

GO ON TO THE NEXT PAGE.

16. How did urban life help acculturate foreign immigrant groups?

    I. The lack of ethnic neighborhoods forced immigrants to adjust to American customs.
    II. Urban political bosses insisted that all remnants of national origin be erased before a person could participate in the electoral process.
    III. English was the language of instruction in the public schools, and the curriculum emphasized American values and institutions.

    F. I only
    G. III only
    H. I and II only
    J. I, II, and III

17. Why did so many nineteenth-century immigrants to the United States settle in cities?

    I. Most were destitute and could not afford to leave the port of entry.
    II. Living costs were lower in the city than in the countryside.
    III. Most were skilled craftsmen, and the demand for their services was greater in the cities.

    A. I only
    B. II only
    C. III only
    D. II and III only

18. Around 1900, blacks thought educational opportunities were greater in northern cities because:

    F. most black colleges were located there.
    G. public education in the North was generally far superior to that in the South.
    H. blacks advocated integration despite separate but truly equal schools in the South.
    J. desegregation of southern schools had reduced the quality of education in that region.

19. Why has it been easier for blacks to be elected to the House of Representatives than to the Senate?

    A. Urban areas tend to reelect the same people to office far more frequently than rural areas, and thus opportunity for new groups is restricted.
    B. A senator is elected to a shorter term than a representative, and thus more businessmen can find time to serve in the Senate.
    C. Senators are elected from a larger geographic area than representatives, and thus the black vote, although smaller than the white vote, frequently holds the balance of power.
    D. Senators are elected on a statewide basis, and this tends to dilute the ability of an urban bloc vote to nominate a minority candidate.

GO ON TO THE NEXT PAGE.

3

Social movements occur when many people, acting together, become aroused about aspects of their social life that they strongly feel need to be changed. Relationships become so unsettled that people have trouble adjusting; disturbing feelings of loneliness result. These feelings develop spontaneously out of widespread conditions of frustration. Individuals attempt to join others like themselves in order to find better solutions to their problems. As they communicate with others, their sense of deprivation, injustice, and impatience with the way problems are being resolved increases.

Social movements are designed to change or to resist change in one or more than one *social institution*—or the system of rules that govern how people act and handle their problems. These rules, which may have become laws, regulate relationships in such areas as the family, government, education, economic activity, and intergroup contact. Many of these relationships have acquired obligatory characteristics and are not easily changed. Strong social pressure is exerted on individuals to continue acting as they have in the past. Efforts aimed originally at changing certain relationships are counteracted and are often redirected later to other goals. The growth of social movements is frequently related to how successfully initial goals of those movements are achieved.

As people increase their participation in social movements, their ties to previous social relationships are weakened. Constant interaction with others provides a sense of togetherness and causes participants to believe they can accomplish almost any feat. Their acts are motivated largely by momentary emotional discomforts rather than by customary social restraints. Traditional ways of behaving gradually lose some of their cohesive power and can be challenged. These new ways of acting are repeated, becoming both the examples for other people's conduct and the bases for new social relationships, laws, and institutions.

Membership in social movements tends to fluctuate during the early periods of their growth, with a small number of people being firmly committed and a much larger number being only marginally involved. The smaller group, which is the core of a social movement, is usually organized around a charismatic leader. The leader personifies the character of the movement and actively seeks to enlarge the membership. Through such activities as rallies, ceremonies, parades, and pamphleteering, members associate with each other and conduct initial recruitment. Later, the media are used to publicize goals, retain old members, and attract new recruits. In time, the organization of social movements becomes formal and established, resembling in many respects the pattern of any other large formal group.

20. Which statement correctly describes the nature of social relationships?

- **F.** Social relationships force people to conform in certain ways.
- **G.** Social relationships seldom pertain to groups other than the family.
- **H.** Social relationships can usually be changed without difficulty.
- **J.** Little or no sentiment is attached to social relationships.

21. The occurrence of which event best indicates the success of a social movement?

- **A.** The formation of a professional football team
- **B.** The American Revolution
- **C.** World War II
- **D.** The selection of a corporation's president

22. Which activity could most likely be associated with a social movement?

- **F.** People leaving a stadium following the defeat of their favorite team
- **G.** The daily movement of people from the city to the suburbs
- **H.** A letter-writing campaign to solicit support for the Equal Rights Amendment
- **J.** A group of people passing out leaflets to advertise a forthcoming movie for a theater chain

23. Meetings conducted by members of social movements often include songs and emotional speeches probably because:

- **A.** every member must be given the chance to express his views fully.
- **B.** it is essential to reinforce a sense of community.
- **C.** leaders can thereby maintain control of their followers.
- **D.** other opportunities for interpersonal communication are unavailable.

24. The author implies that during the early development of social movements, the relationship between the leader and the core membership is:

- **F.** direct and strong.
- **G.** impersonal and weak.
- **H.** unpredictable and callous.
- **J.** accidental and undefined.

25. The people involved in social movements often consider their ideas "antiestablishment" because of:

- **A.** the ineffectiveness of bureaucratic regulations.
- **B.** their perception of weakness in the leadership of social movements.
- **C.** their detachment from existing social constraints.
- **D.** the influence of ideas as sources of social power.

26. According to the passage, which statement best describes the relationship between social movements and social institutions?

- **F.** When social institutions fail, social movements cannot develop.
- **G.** Social institutions and social movements refer to the same kind of behavior.
- **H.** Social movements are more effective when social institutions are strong.
- **J.** Social institutions and social movements have reciprocal influences.

27. Which circumstance is most likely to prevail during the early periods of social movements?

- **A.** Logical discussion of issues
- **B.** Careful attention to democratic procedures
- **C.** Uncompromising commitment to a position
- **D.** Willingness to appreciate all sides of a problem

GO ON TO THE NEXT PAGE.

28. Which phenomenon could most accurately be called a social institution?

F. A classroom activity
G. A longtime friendship
H. The way people get married
J. A clever business practice

Questions 29-43 are not based on a reading passage. You are to answer these questions on the basis of your previous schoolwork in the social studies.

29. The president helps shape public policy when he:

A. delivers the State-of-the-Union address.
B. makes his budget address.
C. issues executive orders.
D. does all of the above.

30. Under *home rule*, a city government can govern its residents with less interference than otherwise from the:

F. county government.
G. president.
H. Supreme Court.
J. state government.

31. Citizens of the United States practice a folkway when they:

A. applaud the performer at a circus.
B. add chlorine to the city water supply.
C. refuse to pay a debt.
D. obey the school-zone speed limit.

32. Regardless of the type of economic system, economic growth can occur ONLY when:

F. values of free enterprise and individual initiative are present.
G. capital gained from productive enterprises is used to improve production.
H. a nation engages in a great deal of foreign trade.
J. the population is growing.

33. Sleep, drug states, and hypnosis are all states of consciousness that can be described as being:

A. abnormal.
B. normal.
C. altered.
D. psychic.

34. The Supreme Court decisions in *Escobedo* v. *Illinois* and *Miranda* v. *Arizona* were concerned with:

F. legislative apportionment.
G. the rights of accused persons.
H. religious freedom.
J. the national government's power of taxation.

35. What is the impact of *third* or *minor* political parties on the American political system?

A. They usually operate outside the legal government structure.
B. They typically influence the major political parties to promote third-party reforms.
C. They generally exercise direct pressure on public officials to oppose any major government changes.
D. They usually prevent the election of major-party candidates.

36. According to the Constitution, the House of Representatives—and NOT the Senate—performs which of the following functions?

I. Approving appointments made [illegible] the president
II. Approving treaties
III. Originating revenue-raising bills

F. I only
G. III only
H. I and II only
J. I, II, and III

37. Which tax generally affects people with lower incomes more severely than it affects people with higher incomes?

A. Estate and inheritance taxes
B. Sales taxes on food and retail items
C. Corporate taxes
D. The progressive federal income tax

38. The nineteenth-century Populists were motivated chiefly by their desire to:

F. achieve justice for the common man.
G. put an end to slavery.
H. maintain an equal number of free and slave states.
J. extend U.S. influence in world affairs.

GO ON TO THE NEXT PAGE.

3

39. The Treaty of Versailles, which concluded World War I, imposed which of the following conditions on Germany?

I. Germany was forced to pay allied war costs.
II. Germany ceded territory to France.
III. Allied forces occupied all of Germany.

A. I only
B. I and II only
C. II and III only
D. I, II, and III

40. The term *Scramble for Africa* describes the period when imperialism was exercised by:

F. Africans, who struggled to regain control over their own destiny.
G. Europeans, who established colonies over most of the continent.
H. Christian denominations, who competed for African converts.
J. Europeans, who competed vigorously with one another for African slaves.

41. A *matrilineal* society is one in which:

A. a woman [illegible] have only one husband at a time.
B. a male marries one of his mother's kin.
C. ancestry is traced through the mother's line.
D. children take both their mother's and father's names as surnames.

42. Which statement best distinguishes between the feudal and the manorial systems of the Middle Ages?

F. Feudalism was based on money, while manorialism was based on land.
G. Feudalism was associated with chivalry, while manorialism was associated with chastity.
H. Feudalism refers to a political and social structure, while manorialism refers to an economic structure.
J. Feudalism refers to an economic structure, while manorialism refers to a political and social structure.

43. Before World War II, arms limitation agreements typically established which kind of restrictions?

A. Quotas on naval armaments
B. Selective limits on military ground forces
C. Courts of arbitration to handle international disagreements
D. Limits on ground-to-air missiles

The following speakers present different viewpoints on the problem of air pollution by automobiles.

Speaker 1

If consumers wanted automobiles with expensive antipollution devices, they could buy such devices as optional equipment. The beauty of American capitalism lies in the fact that the consumer is king. The government should not restrict the freedom of choice of its citizens.

Speaker 2

Consumers are willing to buy only those goods that benefit them directly. When a motorist is driving his automobile, he breathes little, if any, of the pollution emitted by his car. Since he receives little benefit, he will not want to buy the equipment. On the other hand, he would benefit from antipollution equipment on the automobiles of other motorists, but they too are unlikely to spend money for equipment that would benefit someone else. When the benefits of a commodity go to someone other than the buyer, the government must step in to insure that the commodity is produced. Antipollution equipment must be required by law.

Speaker 3

People should be able to buy any type of automobile they desire. Why not measure the amount of pollution produced by every make, assign a monetary value to it, and then impose a federal license fee of an equivalent amount? This would encourage people to choose automobiles with less harmful environmental impact.

Speaker 4

Increased taxes will make automobile ownership more expensive. Fewer people will be able to afford cars. Detroit will get sick, and automotive workers will be standing in front of unemployment offices rather than assembly lines. Perhaps we should give a tax break to those companies that are willing to hire engineers to develop cheaper antipollution equipment.

Speaker 5

Competition in the automotive industry has always led to better products. Power steering, automatic transmission, and other improvements came about without government intervention. I am sure Detroit will eventually improve the environmental impact of its products.

44. Which speaker most clearly expresses the view that people's economic decisions are motivated solely by self-interest?

F. 1
G. 2
H. 3
J. 5

GO ON TO THE NEXT PAGE.

45. Which speaker implies that historical trends can be used to predict the future?
    - A. 1
    - B. 2
    - C. 4
    - D. 5

46. Which government action would Speaker 4 most likely support?
    - F. Passing a law requiring equipment that decreases automobile emissions
    - G. Establishing an agency to study the impact of urban pollution on the health of the citizens
    - H. Imposing a 10% sales tax on automobiles with high emissions
    - J. Awarding a bonus to the company that first produces emission-lowering equipment

47. Which pair of speakers would differ most on the issue of government intervention in the economy?
    - A. 1 and 2
    - B. 1 and 5
    - C. 2 and 3
    - D. 3 and 4

48. When Speaker 4 declares that "Detroit will get sick," he means that:
    - F. smog in Detroit will increase and cause serious respiratory diseases.
    - G. engineers in Detroit will become unemployed.
    - H. automobile sales will decrease throughout the nation.
    - J. Detroit will be forced to default on its municipal bonds.

49. With which statement would Speakers 3 and 4 agree?
    - A. Antipollution equipment makes automobiles more expensive.
    - B. Fighting pollution will increase unemployment.
    - C. Price is the major determinant of consumer spending.
    - D. People will never buy automobiles with antipollution devices.

50. Speaker 3 recommends a tax *increase* while Speaker 4 recommends a tax *decrease* because they disagree about the:
    - F. seriousness of the pollution problem.
    - G. cost of solving the pollution problem.
    - H. effect of increased taxes on the demand for autos.
    - J. willingness of consumers to accept tax cuts.

51. Which government action would reflect Speaker 2's economic philosophy?
    - A. Making seat belts mandatory equipment
    - B. Tying license fees to horsepower
    - C. Raising the federal gasoline tax
    - D. Taxing people who benefit from living near public transportation

52. Some economists believe that large companies decide what they want to produce and then use advertising to persuade consumers to buy the goods. This belief CONTRADICTS the viewpoint of Speaker:
    - F. 1.
    - G. 3.
    - H. 4.
    - J. 5.

END OF TEST 3

STOP! DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

TEST 4

NATURAL SCIENCES READING

35 Minutes—52 Questions

DIRECTIONS: Each passage in this test is followed by several questions. After reading a passage, choose the best answer to each question and blacken the corresponding space on your answer sheet. You may refer to the passage as often as necessary.

Insectivorous plants are highly adapted for capturing and digesting insects through unique structures. Typically green plants, most insectivorous plants manufacture food by photosynthesis and are capable of absorbing nitrogen from the soil through their roots. However, insectivorous plants have evolved adaptations that allow for the digestion and the absorption of nitrogen-rich proteins through the leaf. Organisms that are captured by specialized leaves provide the nitrogen necessary for the plant to survive if the plant lives in a nitrogen-poor habitat.

The pitcher plant is one insectivorous plant that survives in wet, boggy areas where there is little nitrogen available in the soil. This plant, which has hollow leaves shaped like pitchers or horns, uses a pitfall trap to snare organisms. Brightly colored leaves exude a sweet smell to attract the insect, and nectar glands further entice the insect into the mouth of the leaf. Once inside, however, the insect cannot easily escape. Downward-pointing hairs on the inner walls of the pitcher hasten the insect's fall into a cistern containing water and digestive enzymes. Once the insect is digested, nitrogen and other materials are absorbed into the leaf by special cells lining the cistern.

The Venus's-flytrap is another example of an insectivorous plant. Like the pitcher plant, its leaves are adapted for insect trapping. The leaves are hinged in the center and have sharp spines along their outer edges. When an insect brushes at least 2 of the interior trigger hairs, the leaves snap shut like a mousetrap. Once an insect is caught, digestive fluids slowly dissolve its edible parts. After several days, the leaf reopens.

The sundew plant, found in bogs and wet peaty areas, is another type of insectivorous plant. Its method of trapping is similar to flypaper: the ends of the leaves are covered with small tentacles or short hairs which bear a clear, sticky liquid. An insect mistakes this liquid for dew and alights on the gluelike droplet, thus becoming securely trapped. Soon the plant leaf bends, and the surrounding tentacles encompass the insect. In time the insect is digested.

1. One adaptation of the pitcher plant that prevents the escape of the captured insect is the:

   A. sweet-smelling flower.
   B. hollow leaf.
   C. brightly colored leaves.
   D. downward-pointing hairs.

2. Some insectivorous plants might be sweet-smelling while others might have a pungent odor because the plants have:

   F. reached different evolutionary stages.
   G. adapted to different habitats.
   H. adapted to attract different organisms.
   J. different flower colors.

3. The pitcher plant survives in areas where most other noninsectivorous plants do not because the:

   A. seeds of most other plants do not fall in the bog.
   B. digestive enzymes of the pitcher plant kill most other plants.
   C. lack of nitrogen in the soil prevents the growth of most other plants.
   D. insects of the area consume most other plants.

4. Which factor was probably LEAST relevant to the development of the trapping mechanisms of the sundew plant?

   F. Favorable mutations of the plant's genetic code
   G. Types of organisms captured by the plant
   H. Lack of nitrogen in the boggy soil
   J. Color of the plant's flower

5. An experimenter fails to stimulate a sundew plant with sugar, starch, and alcohol solutions but is successful with solutions containing milk, albumen, and ammonium salts. The experimenter might conclude that the sundew plant:

   A. selectively chooses its nutrient sources.
   B. responds more readily to albumen than to milk.
   C. is sensitive to some touch stimuli.
   D. can digest food in solute form more easily than it can insects.

6. The digestive enzyme secreted by insectivorous plants would be expected to be analogous to that enzyme found in humans called:

   F. pepsin.
   G. carboxylase.
   H. catalase.
   J. lactase.

GO ON TO THE NEXT PAGE.

7. If 2 trigger hairs in the Venus's-flytrap are stimulated, or if 1 trigger hair is stimulated twice, the leaf will close. This adaptation has occurred in order that:

   A. the leaves do not spring shut for false alarms.
   B. an animal too large for the flytrap can escape.
   C. the leaves close when brushed by any foreign material.
   D. the plant can chemically determine if the organism contains nitrogen.

8. The prickles of a thistle have evolved in order that the thistle can be carried by:

   F. wind.
   G. grazing animals.
   H. birds.
   J. bees.

9. Which is NOT an example of plant adaptation?

   A. Thick bark of an oak tree
   B. Shallow, widespread root system of a cactus
   C. Red leaves of an autumnal maple tree
   D. Needle-type leaf of a pine tree

10. A Venus's-flytrap plant is raised under normal greenhouse conditions and is not fed any insects. Does this plant survive?

   F. Yes, because adequate nitrogen is available in the soil.
   G. Yes, because the plant does not need to compete with other plants for its nutrients.
   H. No, because only insects provide the nutrients necessary to sustain the plant.
   J. No, because the plant cannot adjust to the greenhouse environment.

GO ON TO THE NEXT PAGE.

4

To understand why objects float or sink when placed in water, one must study the general relationship between the forces acting on an object immersed in water and the motion of that object in water.

Newton's second law defines the general relationship between forces and motion as

$$a = \frac{F}{m}$$

where a = the acceleration of the object,
F = the net force acting on the object,
and m = the mass of the object.

An object at rest or moving in a straight line with a constant velocity has *zero* acceleration. Hence, the net force acting on it must also be *zero*. If the net force acting on an object is *zero*, then the sum of the forces acting in one direction must equal the sum of the forces acting in the opposite direction.

When an object is in water, however, there may be several forces acting on it. The earth will exert a gravitational force on the object; this downward force is called the weight of the object. Buoyant force acts on an object immersed in water as well. Archimedes' principle states that a body immersed in a liquid is buoyed up with a force equal to the weight of the displaced liquid.

A student conducted 3 experiments to validate Archimedes' principle. The same steel block was used in all the experiments.

Experiment 1

To determine the weight of the steel block, the student hung the block from a spring balance. The balance read 9.5 newtons.

Experiment 2

To determine the force of the water on the block, the student again hung the block from the balance. The block (but not the balance) was then submerged in a tank of water. The block did not touch the bottom or sides of the tank. The balance read 8.3 newtons.

Experiment 3

To determine the weight of the water displaced by the block, the student put a pan weighing 5.7 newtons under the tank. The tank was then filled to the brim with water, the block was submerged in the tank, and the water that overflowed the brim was caught in the pan. The weight of the pan plus the water that it caught was 6.8 newtons.

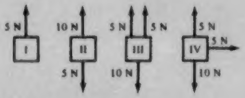
11. In Experiment 3, the student filled the tank to the brim in order to:

- **A.** help determine the initial amount of water.
- **B.** make sufficient room to submerge the block completely.
- **C.** ensure an accurate measure of the amount of water displaced by the block.
- **D.** maximize the buoyant force of the water on the block.

12. When an object is placed in water, the buoyant force of the water on the object is equal to:

- **F.** the weight of the object.
- **G.** the weight of the water displaced by the object when it enters the water.
- **H.** the mass of the object multiplied by its acceleration.
- **J.** zero.

13. In the figure below, the directions of the forces acting on each of 4 objects (labeled I through IV) are indicated by arrows, and the magnitudes of the forces (in newtons) are noted beside the arrows. Which object has *zero* net force acting on it?



- **A.** I
- **B.** II
- **C.** III
- **D.** IV

14. Why did the readings on the spring balance in Experiments 1 and 2 differ?

- **F.** The water does not exert a force on the block in Experiment 1 but does in Experiment 2.
- **G.** The acceleration of the block was equal to zero in the second experiment.
- **H.** The force of the earth on the block in Experiment 1 is greater than in Experiment 2.
- **J.** The force of the earth on the block in Experiment 1 is less than in Experiment 2.

15. In the design of which of the following would builders be LEAST likely to use Archimedes' principle?

- **A.** Canoes
- **B.** Weather balloons
- **C.** Submarines
- **D.** Rockets

16. On the basis of Experiment 3, what is the weight of the water, in newtons, displaced by the block?

- **F.** 0.0
- **G.** 1.1
- **H.** 5.7
- **J.** 6.8

GO ON TO THE NEXT PAGE.

17. In Experiment 2, what is the buoyant force of the water, in newtons, on the block?

- **A.** 0.0
- **B.** 1.2
- **C.** 8.3
- **D.** 9.5

Use the following information to answer questions 18 and 19.

The table below lists the masses and net forces acting on 5 objects.

Object	Mass (kg)	Net Force (N)
I	1	10
II	5	20
III	10	20
IV	40	40
V	20	80

18. Which of the 5 objects would have the greatest acceleration?

- **F.** I
- **G.** III
- **H.** IV
- **J.** V

19. Suppose the 5 objects all had a net force of 10 newtons acting upon them. Which object would have the greatest acceleration?

- **A.** I
- **B.** II
- **C.** III
- **D.** IV

Questions 20-34 are not based on a reading passage. You are to answer these questions on the basis of your previous schoolwork in the natural sciences.

20. To achieve the hottest flame possible using a Bunsen burner, one should adjust the burner so that the flame is:

- **F.** as small as possible.
- **G.** bright yellow.
- **H.** pale blue.
- **J.** pale blue with a darker blue inner core.

21. Which material does NOT act as a conductor?

- **A.** Copper wire
- **B.** Silver metal
- **C.** Salt water
- **D.** Glass tubing

22. A person hypothesizes that a room can be cooled if an electric refrigerator, which stands in the middle of the room, is left open while it operates. After a long period of time, however, she observes that the temperature of the room is increasing rather than decreasing. Which statement explains this occurrence?

- **F.** The gas that leaks from the refrigerator releases heat as it expands.
- **G.** The freezing compartment is not open.
- **H.** The refrigerator motor generates heat.
- **J.** The refrigerator is not powerful enough to cool the room.

23. If a man who is left-handed and blue-eyed (both recessive traits) marries a woman who is left-handed and brown-eyed, their children could inherit which traits?

I. Left-handedness and blue eyes
II. Left-handedness and brown eyes
III. Right-handedness and blue eyes

- **A.** I only
- **B.** II only
- **C.** I or II only
- **D.** I or III only

24. Dew, which is essentially moisture condensed out of the air, will form when there is:

- **F.** a decrease in temperature.
- **G.** an increase in barometric pressure.
- **H.** a decrease in barometric pressure.
- **J.** an absence of wind.

25. Botanists define a fruit as a ripened ovary. According to this definition, which food is a fruit?

- **A.** Sweet potato
- **B.** Turnip
- **C.** Cucumber
- **D.** Beet

26. Air bubbles generated deep under water increase in volume as they rise to the surface because the:

- **F.** pressure inside the bubbles increases.
- **G.** air inside the bubbles decreases in temperature.
- **H.** air in the bubbles dissolves in the water.
- **J.** pressure of the water on the bubbles decreases.

27. Since molds cannot conduct photosynthesis, it must be true that in their interaction with their environment they:

- **A.** must inhabit areas receiving 8 hours of sunlight in order to flourish.
- **B.** must obtain organic nutrients from their environment.
- **C.** deplete the oxygen in their environment, and therefore their growth is limited.
- **D.** tend to lose organic nutrients to their environment.

GO ON TO THE NEXT PAGE.

28. The kinetic theory of gases assumes that there are no attractive forces between the molecules of an ideal gas. Which observation supports this assumption?

F. Gases at ordinary temperatures and pressures are highly compressible.
G. Gases tend to liquefy at high pressures and at cool temperatures.
H. The molar volume of all gases at standard temperature and pressure is approximately 22.4 liters.
J. Gases spontaneously expand to occupy all the volume available to them.

29. To determine whether kerosene is an electrolyte in water, one should:

A. measure its solubility in water.
B. measure the density of the kerosene-and-water solution.
C. measure its contribution to the conductivity of water.
D. observe its electronic spectra in water with a spectroscope.

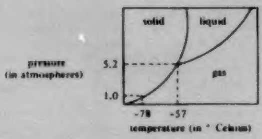
30. In which property do X rays and light rays differ?

F. Wavelength
G. Speed
H. Charge
J. None of the above

31. The genes in a cell would be immediately altered if the experimenter substituted in the cell one:

A. amino acid for another in a protein.
B. nucleotide for another in DNA.
C. nucleotide for another in RNA.
D. sugar for another in a starch.

32. What does the phase diagram of carbon dioxide shown below reveal about carbon dioxide?



F. Tremendous pressure is needed to liquefy carbon dioxide at -78° C.
G. Carbon dioxide cannot exist as a liquid at normal atmospheric pressure.
H. Carbon dioxide cannot exist as a gas below -78° C.
J. The triple point of carbon dioxide occurs at -78° C and 1.0 atm pressure.

33. The earth's equatorial radius is greater than its polar radius due to the:

A. earth's rotation.
B. gravitational attraction of the sun.
C. gravitational attraction of the moon.
D. higher temperatures near the equator.

34. Adenosine triphosphate (ATP) is important to the active transport system of the cell membrane because it provides the cell membrane with the necessary:

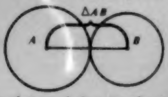
F. fluid.
G. energy.
H. isotopes.
J. chlorophyll.

Theoretical and physical molecular models, which show how electrons form the bonds that hold molecules together, are used by chemists to illustrate relationships between atoms in a molecule. One theoretical molecular model places electrons in concentric orbits around a central atom; each orbit or *shell* has a maximum capacity of 8 electrons. A full outer shell is regarded as the most stable electronic configuration that an atom can have. A bonding theory introduced by Lewis uses this principle. Lewis assumes that (1) a bond between atoms involves 2 electrons that are shared by the atoms they join, and (2) each atom is surrounded by 4 pairs, or 8, electrons. Exceptions to this *octet theory* occur in the cases of the lightest elements. Hydrogen needs only 1 pair of electrons for a complete outer shell. Beryllium, boron, and aluminum sometimes form compounds with, respectively, 2, 3, and 3 pairs in their outer shells.

Structural models can illustrate the octet theory of Lewis. Because only the electrons in the outer shell of an atom come in contact when 2 atoms bond together, only the outer electrons are depicted in the model. The atoms' electrons are arranged to give each atom in the molecule a group of 8 electrons. The following example shows how the individual outer electrons, represented by dots, bond in carbon dioxide; the short lines represent the electron pairs in the molecule.

$$:\ddot{O}::C::\ddot{O}: \qquad |\overline{O}=C=\overline{O}|$$

To translate these theoretical structural models into physical models, the chemist determines the size and shape of the molecule. The size of each atom in the molecule can be estimated by calculating the atom's radius from (1) the experimentally determined bond length between bonding atoms *A* and *B* (ΔAB), and (2) the electronegativities of the 2 atoms.



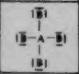
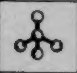
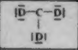
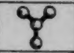
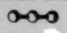
The radius (*r*) of atom *A* then may be calculated by the formula:

$$r_A = \frac{\text{electronegativity}_A}{\text{electronegativity}_A + \text{electronegativity}_B} \times \begin{array}{c}\text{bond}\\ \text{length}\\ \text{between}\\ A \text{ and } B\\ (\Delta AB)\end{array}$$

The radius of *B* is then equal to the bond length minus the radius of *A*.

GO ON TO THE NEXT PAGE.

The molecule's shape is determined by the number of sets of electron pairs surrounding the central atom. Typically, the sets are arranged in space so that they are separated by the greatest distance and angle possible. The table below illustrates this concept.

Geometry	Structural Model	Physical Model
Tetrahedral: 4 sets of electron pairs surrounding the central atom		
Trigonal planar: 3 sets of electron pairs surrounding the central atom		
Linear: 2 sets of electron pairs surrounding the central atom	$\overline{F}=E=\overline{F}$	

35. The primary factor that determines the shape of a molecule is the:
    - **A.** number of sets of electron pairs surrounding the molecule's central atom.
    - **B.** bond distance between 2 atoms in the molecule.
    - **C.** total number of electrons in the molecule.
    - **D.** length of the radii of individual atoms in the molecule.

36. If the central atom of a molecule were surrounded by 3 sets of electron pairs, the geometry of the molecule would probably be:

    I. linear.
    II. trigonal planar.
    III. tetrahedral.

    - **F.** I only
    - **G.** II only
    - **H.** I or III only
    - **J.** II or III only

37. Hydrogen (H) has 1 electron in its outer shell, and nitrogen (N) has 5. Which symbol accurately depicts the electron-dot structure for NH_3?
    - **A.** N:H:H:H
    - **B.** H: :N:H
H
    - **C.** N:H:H
    - **D.** H:N:H
H

38. The structural model for carbon dioxide depicted in the second paragraph suggests that:
    - **F.** a double bond can be formed between 2 atoms.
    - **G.** 8 electrons form 8 bonds.
    - **H.** carbon dioxide is not a stable molecule.
    - **J.** only half of carbon's outer shell electrons are used to form the bonds with oxygen.

39. To calculate the size of an atom in a molecule, one needs to know the:

    I. electronegativity of that atom and of the one to which it is bonded.
    II. bond distance from that atom to the one to which it is bonded.
    III. number of sets of electron pairs surrounding the central atom.
    IV. total number of electrons surrounding that atom.

    - **A.** II only
    - **B.** III only
    - **C.** I and II only
    - **D.** III and IV only

40. According to the Lewis theory, the electrons that are involved in bonding must be:
    - **F.** contributed by both bonding atoms.
    - **G.** transferred from one atom to the other.
    - **H.** shared by both atoms, regardless of the electrons' origin.
    - **J.** paired with 4 other electrons to form a bonding octet.

41. Sets of electrons are usually found separated by the greatest possible distance and angle because:
    - **A.** electron shells can be broken only into equal parts.
    - **B.** atoms are most stable at the corners of regular geometric shapes.
    - **C.** the central atom attracts the atoms' electron pairs.
    - **D.** negative charges cause the electrons to repel each other.

42. The bond length between the hydrogen and carbon atoms in CH_4 is 1.09 angstroms, and the electronegativity for carbon is 2.5 and that for hydrogen is 2.1. Which equation should be used to figure the radius of the carbon atom?
    - **F.** $r = 2.5 \times 1.09$
    - **G.** $r = 2.5 \times 2.1 \times 1.09$
    - **H.** $r = (2.5 + 2.1) \times \frac{1.09}{2.5}$
    - **J.** $r = \frac{2.5}{2.5 + 2.1} \times 1.09$

43. Hydrogen needs only 1 pair of electrons for a complete outer shell. Which statement is INCONSISTENT with this model of hydrogen?
    - **A.** Hydrogen atoms are not found at the centers of molecules.
    - **B.** Hydrogen atoms do not determine the geometries of most molecules.
    - **C.** An outer shell containing 8 electrons is required for a stable configuration.
    - **D.** Only the outer shell electrons are involved in bonding.

GO ON TO THE NEXT PAGE.

4

Bordering the continents is an undersea ledge of rock called the *continental shelf*. At its edge, the *continental slope* begins; the continental slope drops to the deep-ocean floor below. Much of the deep-ocean floor is rugged, covered with ridges and trenches. Only the *abyssal plains*, scattered below and beyond the continental slope, are flat and smooth. Though their rocky bottoms are as rugged as the rest of the seafloor, they are covered with great thicknesses of sediments—sand, silt, and clay.

How did these sediments come to the abyssal plains? A study of a 1929 earthquake that originated on Newfoundland's continental slope helped provide information regarding the origin of these sediments. After the earthquake, undersea telegraph cables snapped. If the earthquake itself had been responsible for the breakage, all the cables should have broken at nearly the same instant. The data in the table below, however, reveal otherwise.

Cable #	Distance South of Edge of Continental Shelf (in kilometers)	Time of Break After Earthquake
1-8	all within 160	instantaneous
9	240	59 min.
10	450	3 hr. 3 min.
11	730	9 hr. 1 min.
12	780	10 hr. 18 min.
13	800	13 hr. 17 min.

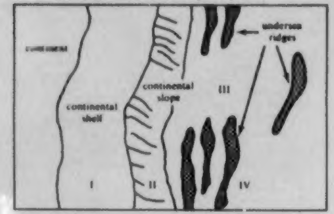
Scientists hypothesized that the earthquake had started an undersea landslide which broke the first 8 cables. Sediments then filled the water, increasing its density. This much denser water began to flow down the slope, gathering more sediments and increasing its speed, and snapped the cables in its path. Reaching the flat floor of the abyssal plains, the sediment-dense water slowed. Finally, the sediments settled out of the water, forming a layer with the coarser materials at the bottom and the finer materials at the top.

Columnar samples of sediments in the area revealed a graded, 1-meter layer of sediments that contained the remains of continental-shelf organisms. Thus, the scientists concluded, these flows of sediment-laden water, called *turbidity currents*, must have created the abyssal plains. Such currents need only a source of sediments, some kind of disturbance (such as a storm or an earthquake) to set them in motion, and an unobstructed path to the seafloor.

Turbidity currents also account for submarine canyons. For years, scientists believed that these canyons were originally above sea level on the continental margins and had been formed by terrestrial rivers. Now they have concluded that turbidity currents must have carved them out.

44. Sediments that cover the abyssal plains are carried there from:
    - **F.** deep-ocean trenches.
    - **G.** the continental shelf.
    - **H.** submarine canyons.
    - **J.** coastal rivers.

45. In the diagram below, which numeral indicates the most likely location of an abyssal plain?



    - **A.** I
    - **B.** II
    - **C.** III
    - **D.** IV

46. Compared to the organisms living on the abyssal plains, those creatures inhabiting the continental shelves are best adapted to a habitat which is:
    - **F.** shallower.
    - **G.** colder.
    - **H.** smoother.
    - **J.** darker.

47. Suppose a certain submarine canyon becomes deeper after storms and shallower during quiet weather. This observation would offer support for the hypothesis that storms:
    - **A.** cause weather that resembles earthquake activity.
    - **B.** set off landslides on the continental shelf that fill submarine canyons with sediments.
    - **C.** set off turbidity currents in submarine canyons that carry accumulated sediments beyond the continental slope.
    - **D.** set off turbidity currents in submarine canyons that carry accumulated sediments onto the continental slope.

48. Which statement supports the scientists' hypothesis that the first 8 cables were broken by the landslide generated by the Newfoundland earthquake?
    - **F.** The cables were within 160 kilometers of the edge of the continental shelf.
    - **G.** The cables broke almost at the instant the earthquake occurred.
    - **H.** The cables were located quite close to one another.
    - **J.** Turbidity currents need an unobstructed path if they are to stay in motion.

GO ON TO THE NEXT PAGE.

49. What should scientists do to determine how many turbidity-current events produced the sediments accumulated over a particular abyssal plain? (Assume all measurements described can be made.)
    - A. Count the number of layers of sediment accumulated on the abyssal plain.
    - B. Tally how many years of sediments have accumulated, assuming that 1 meter of sediment represents 1 year's accumulation.
    - C. Count the number of earthquakes and storms in the area each year and multiply this number by the age of the earth.
    - D. Measure the thickness of sediments settling out of seawater in 1 year and divide this number into the total thickness of the sediments overlying the abyssal plain.

50. Which of the following observations suggested to the scientists that a force other than an earthquake had broken the cables off the Newfoundland shore?
    - F. A great number of cables were broken.
    - G. Great force was required to snap an undersea cable.
    - H. The cables broke at widely varying times.
    - J. All of the above

51. If the columnar samples of the seafloor off Newfoundland's continental slope had revealed nothing but rocky seafloor, the scientists would most likely have concluded that the cable breaks could be accounted for by:
    - A. a turbidity current that had been set off by some event other than an earthquake.
    - B. the landslide.
    - C. some event other than a turbidity current.
    - D. a turbidity current that had carried no sediments.

52. The information in the passage would support which of the following statements about turbidity currents?
    - F. They travel at an average speed of 200 kilometers per hour.
    - G. They are agents of deposition, but not of erosion.
    - H. They may flow for hours before the sediments in the currents settle out.
    - J. All of the above

END OF TEST 4

STOP! DO NOT RETURN TO ANY OTHER TEST.

Answer Key to Sample Test Booklet

Test 1 - English Usage

1 C	11 C	21 A	31 B	41 D	51 A	61 D	71 D
2 F	12 G	22 J	32 H	42 G	52 F	62 G	72 F
3 B	13 D	23 D	33 D	43 B	53 A	63 A	73 D
4 J	14 H	24 J	34 F	44 J	54 H	64 J	74 H
5 A	15 D	25 C	35 D	45 D	55 D	65 B	75 C
6 H	16 H	26 J	36 J	46 F	56 J	66 G	
7 C	17 B	27 B	37 D	47 D	57 D	67 A	
8 J	18 F	28 H	38 G	48 F	58 G	68 F	
9 A	19 A	29 A	39 D	49 D	59 C	69 B	
10 G	20 F	30 F	40 G	50 J	60 H	70 J	

Test 2 - Mathematics Usage

1. D	11. C	21. C	31. B
2. H	12. G	22. J	32. H
3. C	13. C	23. D	33. D
4. F	14. F	24. G	34. K
5. C	15. A	25. C	35. A
6. K	16. G	26. J	36. K
7. D	17. C	27. C	37. D
8. F	18. G	28. J	38. J
9. C	19. C	29. C	39. B
10. K	20. H	30. G	40. F

Test 3 - Social Studies Reading

1 A	11 D	21 B	31 A	41 C	51 A
2 H	12 G	22 H	32 G	42 H	52 F
3 D	13 B	23 B	33 C	43 A	
4 H	14 F	24 F	34 G	44 G	
5 D	15 B	25 C	35 B	45 D	
6 H	16 G	26 J	36 G	46 J	
7 D	17 A	27 C	37 B	47 A	
8 F	18 G	28 H	38 F	48 H	
9 B	19 D	29 D	39 B	49 C	
10 J	20 F	30 J	40 G	50 H	

Test 4 - Natural Sciences Reading

1 D	11 C	21 D	31 B	41 D	51 C
2 H	12 G	22 H	32 G	42 J	52 H
3 C	13 C	23 C	33 A	43 C	
4 J	14 F	24 F	34 G	44 G	
5 A	15 D	25 C	35 A	45 C	
6 F	16 G	26 J	36 G	46 F	
7 A	17 B	27 B	37 D	47 C	
8 G	18 F	28 J	38 F	48 G	
9 C	19 A	29 C	39 C	49 A	
10 F	20 J	30 F	40 H	50 H	

Actual Test Directions

This booklet contains four tests—*English Usage, Mathematics Usage, Social Studies Reading,* and *Natural Sciences Reading.* The four tests measure skills and abilities that are highly related to success in college.

During the first half of the testing session, you will complete the English Usage Test and the Mathematics Usage Test. In the second half, you will complete the Social Studies Reading Test and the Natural Sciences Reading Test.

The questions in each test are numbered, and the four or five suggested answers for each question are lettered. On the answer sheet, the rows of ovals are numbered to match the questions and the ovals in each row are lettered to correspond to the suggested answers.

For each question, first decide which answer is best. Next, locate on the answer sheet the row of ovals numbered the same as the question. Then, locate the oval in that row lettered the same as your answer. Finally, blacken the oval completely. Use a soft lead pencil and make your marks heavy and black. *DO NOT USE A BALL-POINT PEN.*

You may work on each test ONLY when your test administrator tells you to do so. If you finish a test before time is called, you should use the time remaining to reconsider questions you are uncertain about in that test. You may NOT look back to a test on which time has already been called, and you may NOT go ahead to another test. To do so will disqualify you from the examination.

Your score on each test will be based only on the number of questions you answer correctly. You will NOT be penalized for guessing. *HENCE IT IS TO YOUR ADVANTAGE TO ANSWER EVERY QUESTION.*

If you change your mind about an answer, erase your first mark thoroughly before marking your new answer. For each question, make certain that you mark in the row of ovals with the same number as the question.

Do not fold or tear the pages of your test booklet. Keep the booklet in as good a condition as possible.

DO NOT OPEN THIS BOOKLET UNTIL TOLD TO DO SO.

ACT

RESEARCH REPORTS

COMPARATIVE PREDICTIVE VALIDITIES OF THE AMERICAN COLLEGE TESTS AND TWO OTHER SCHOLASTIC APTITUDE TESTS

August, 1965 No. 6

Leo Munday

Published by Research and Development Division · American College Testing Program · P. O. Box 168, Iowa City, Iowa 52240

Summary

This investigation compared the predictive validity of the ACT tests with the SAT and SCAT tests for a sample of 21 colleges and universities. Grades in specific courses as well as over-all grade point average were studied. Multiple R's were computed for each test battery. The ACT and SAT tests possessed about the same degree of predictive validity, and both were better predictors than the SCAT.

Predictive results varied from school to school and from course subject to course subject, thereby suggesting that predictive validity for individual colleges and universities and for specific subject matter areas should be established. In addition, a college's choice between the ACT or SAT probably can not be made on the basis of their relative predictive validities, but rather should be made on the basis of the total programs of services offered by the respective testing agencies.

Comparative Predictive Validities of the American College Tests and Two Other Scholastic Aptitude Tests

Leo Munday

Standardized tests of academic potential are used in nearly all American colleges and universities for admissions and guidance purposes. Institutions have a wide variety of such tests from which to choose. While many considerations are relevant to this choice, a major consideration is predictive validity.

This report compares the predictive validities of three of the most widely used tests of scholastic potential. These were the American College Testing Program's tests (ACT), the College Entrance Examination Board's Scholastic Aptitude Test (SAT), and the Educational Testing Service's School and College Ability Test (SCAT). The ACT battery consists of tests in English, mathematics, social studies, and natural science; it requires 180 minutes. Both the SAT and the SCAT have verbal and quantitative parts; the SAT requires 180 minutes and the SCAT requires 70 minutes.

Procedure

Sample. Both ACT and SAT scores were available for freshmen at five colleges and universities which participated in the 1964 ACT Research Service. ACT and SCAT scores were also available for students at six other colleges and universities.

Additional data from three other studies were made available to the investigator.[1] In two of these (Chase, et al., 1963, Klingelhofer, 1964), the ACT and SAT tests were administered to randomly selected portions of the entering class. SCAT scores were also available for virtually all of Klingelhofer's students. In the other study (Eells, 1962), both ACT and SCAT were administered at eight colleges; and, at two of these, SAT results were available for a segment of the same groups. Altogether, data were available for students at 21 colleges and universities.

Since some of the colleges subdivided their samples by sex or curriculum, we obtained a total of 29 groups from the 21 colleges. All tests were taken prior to college instruction, and all grades were earned during the freshman year. Table 1 describes these institutions in terms of their enrollment, control, level of instruction, and geographic region.

Table 1

Selected Institutional Characteristics of Schools in This Sample

College	Total Enrollment	Level of Instruction	Source of Control	Geographic Region
1	3,910	Offers M. A. Degree	State	West Coast
2	19,557	Offers Ph. D. Degree	State	Rocky Mtn.
3	930	Offers M. A. Degree	State	New England
4	2,040	Four years-liberal arts	Private	New England
5	280	Junior College	Church	Middle West
6	580	Four years-liberal arts	Church	Middle West
7	614	Junior College	Private	New England
8	1,423	Four years	State	South
9	954	Offers M. A. Degree	Church	Southwest
10	31,581	Offers Ph. D. Degree	State	Middle West
11	2,488	Offers Ph. D. Degree	Private	Canada

Table 1 (cont.)

College	Total Enrollment	Level of Instruction	Source of Control	Geographic Region
12	33,956	Offers Ph.D. Degree	State	Middle West
13	5,374	Offers M.A. Degree	State	Middle West
14	3,664	Offers M.A. Degree	State	Middle West
15	6,571	Offers M.A. Degree	State	Middle West
16	9,863	Offers Ph.D. Degree	State	Middle West
17	16,843	Offers Ph.D. Degree	State	Middle West
18	4,624	Offers M.A. Degree	State	Middle West
19	4,146	Offers M.A. Degree	State	Middle West
20	17,024	Offers M.A. Degree	State	West Coast
21	354	Junior College	Community	Middle Atlantic

Criteria. Over-all freshman GPA was available for all but one sample. Additional criteria include GPA's in the following specific freshman courses: English, mathematics, social science, humanities, and science.

Statistical Treatment. Multiple correlations (R) were computed between the four ACT scores and each criterion. The same procedure was followed for the two SAT scores and for the two SCAT scores.

In the application of test results, scores are usually combined with high school grades. However, high school grades were not routinely available except for colleges and universities in the ACT Program. We assumed that all tests would have similar correlations with high school grades, and therefore the increase in validity obtained by adding high school grades to test scores would be relatively constant. This assumption was not formally tested, but it is consistent with most of the literature in academic prediction and with the experience of ACT's Research Service.[2]

Within each criterion area multiple correlations (R) were averaged using Fisher's z procedure (Edwards, 1950). This procedure was not strictly appropriate,[3] but it provided a convenient summary of two columns of R's that took cognizance of relative sample sizes.

Results and Discussion

Table 2 compares the validity of ACT and SAT for predicting over-all grades.

Table 2

Comparative Validity of ACT and SAT

for Predicting Over-all Freshman Grades

College	Group	N*	Multiple Correlation ACT	 SAT
1	Men	308	.33	.38
2	Men	445	.37	.33
20	Men	175 - 183	.47	.38
1	Women	469	.46	.42
2	Women	700	.50	.42
20	Women	223 - 232	.54	.45
3	All Freshmen	127	.45	.44
12	All Freshmen	433	.48	.48
13	All Freshmen	531	.44	.41
1	Vocational Majors	202	.37	.33
4	Business Majors	264	.41	.41
4	Secretarial Educ.	215	.38	.64
11	Science Majors	187	.43	.40
11	Humanities Majors	166	.37	.38
11	Soc. Sci. Majors	149	.41	.43
	Totals	4594 - 4611	.44**	.42**

*When two N's are given, a different sample took each test; ACT N is given first.

**Computed by transforming R's to z's, weighting by N, and retransforming to R's. (Edwards, 1950).

Table 3 summarizes these data and predictive validities in other academic areas for these two batteries.[4] A similar summary of ACT-SCAT comparative validity is found in Table 4.

Table 3

Comparative Validity of ACT and SAT for Predicting College Freshman Grades

	Criterion Area				
	Over-all	English	Math	Soc. Sci.	Science
No. of Colleges	8	5	4	5	4
No. of Correlations	15	10	6	10	11
Range of N's	127 to 700	123 to 585	30 to 605	97 to 261	62 to 254
Total N	4594-4611	2427-1988	1033-1406	1452-1461	1576-1524
Range of R's, ACT	.33 to .54	.27 to .61	.40 to .66	.23 to .70	.31 to .72
Range of R's, SAT	.33 to .64	.29 to .67	.35 to .66	.23 to .66	.30 to .70
No. times ACT higher	9	6	5	7	9
No. times SAT higher	5	4	1	2	1
Mean R*, ACT	.44	.45	.51	.48	.52
Mean R*, SAT	.42	.43	.49	.44	.50

* Computed by transforming R's to z's, weighting by N, and retransforming to R's. (Edwards, 1950). At one institution where sample sizes for some criteria were not available, the average sample size for that school was used.

Table 4

Comparative Validity of ACT and SCAT for Predicting College Freshman Grades

	Criterion Area			
	Over-all	English	S. Sci. or Hum.	Sci. or Math
No. of Colleges	15	5	5	3
No. of Correlations	17	6	6	5
Range of N's	66 to 1857	108 to 255	135 to 227	108 to 204
Total N	9376-9771	993	987	765
Range of R's, ACT	.30 to .69	.51 to .72	.35 to .64	.36 to .54
Range of R's, SCAT	.25 to .70	.41 to .74	.23 to .62	.24 to .58
No. times ACT higher	16	5	5	3
No. times SCAT higher	1	1	0	2
Mean R*, ACT	.53	.57	.53	.44
Mean R*, SCAT	.49	.51	.44	.40

* Computed by transforming R's to z's, weighting by N, and retransforming to R's. (Edwards, 1950)

These tables suggest several conclusions:

1. Predictive results varied widely from one campus to the next. The specific validity of any test should be established for the individual college.[5]

2. Differences were noted in the predictability of curricular and sex subgroups at individual colleges. These results imply that complex colleges should develop several different equations for predicting grades so that various contingencies can be taken into account.

3. The ACT and SAT tests possessed about the same degree of predictive validity. The slight predictive advantage to ACT shown in Table 3 is too small to be of any practical consequence.

4. ACT appears to be a reliably better predictor than SCAT, both in terms of over-all grades and grades in specific courses.

There are, however, certain limitations inherent in these conclusions. In the first place, because the 21 colleges were not selected to be representative of any known population, the statistics can be applied in a strict sense only to the sample institutions. Since a reasonably diverse group of institutions was involved, however, and since the results from college to college were similar, this limitation is probably not serious.

Secondly, we did not know whether one or the other test had been used for admissions purposes. Whatever the situation, the test used for admissions would be at a disadvantage.

And finally, while sample sizes for specific courses were

719

occasionally so small that these samples probably produced unstable correlations, the average R's reported in Tables 2, 3, and 4 are exceptions to this deficiency.

Within these limitations, we can safely conclude that all three tests possess useful predictive validity. SCAT scores were least predictive, probably because SCAT is a much shorter test than the other two. The slight predictive advantage enjoyed by the ACT battery is probably a function of its more comprehensive nature (four tests as opposed to two) and its content (which is explicitly related to typical college freshman courses).

Most colleges today participate in one of the two national testing programs -- ACT and CEEB -- whose tests were compared in this study. Other colleges will likely make a choice soon. Our findings suggest that differences in validity between ACT and SAT were so slight that this choice might better be made on the basis of such other considerations as availability of scores, type of reports on individual students, costs, convenience to the student, and special services offered by the testing agencies.

References

ACT Technical Report. Iowa City: American College Testing Program, 1965.

Chase, C., Ludlow, H., Pomeroy, Martha C., & Barritt, L. Indiana University studies in prediction: II. predicting individual course success for entering freshmen. Bureau of Educational Studies and Testing, Bloomington, Indiana: Indiana University, 1963.

Edwards, A. L. Experimental design in psychological research. New York: Rinehart, 1950.

Eells, K. How will the ACT tests serve Illinois needs? - A Research Report. Committee on Freshman Testing of the Illinois Joint Council on Higher Education, 1962.

Klingelhofer, E. Interrelationships of high school grades, college grade point average and scores on SCAT, ACT, and SAT for a class of college freshmen. Paper presented at the 42nd Annual Conference of the California Educational Research Association, Santa Barbara, California, March 14, 1964.

Footnotes

[1] Grateful acknowledgment is hereby made to Dr. Kenneth Eells, Dr. E. L. Klingelhofer, and Dr. Clinton Chase for their cooperation.

[2] An illustration of the importance of adding high school grades to test score data is found in the ACT Technical Report (American College Testing Program, 1965). For almost 600 colleges, the median correlation of ACT scores with college freshman GPA was .53; when high school grades are added, this correlation is increased to .64.

[3] The assumption that all samples came from the same population was not thought to be tenable.

[4] Tables which give individual correlations for each group and each criterion are in the appendix.

[5] Colleges participating in either the American College Testing Program or the College Entrance Examination Board Program are eligible to receive local validation studies at no cost as an aspect of these programs.

Appendix 1

Comparison of ACT and SAT

The Prediction of College English Course Grades

College	Sample	N*	ACT-R	SAT-R
1	Men	211	.27	.41
1	Women	369	.40	.29
1	Vocational Students	142	.39	.36
3	Freshmen	123	.38	.33
4	Business Majors	257	.46	.34
4	Teacher and Secr'l. Educ.	214	.42	.67
4	Teacher and Secr'l. Educ.	215	.28	.49
10	Freshmen	585 - NA	.61	.55
10	Freshmen	NA - NA	.56	.50
11	Humanities Majors	165	.39	.40

The Prediction of College Mathematics Course Grades

College	Sample	N*	ACT-R	SAT-R
3	Freshmen	103	.51	.46
4	Business Majors	259	.62	.57
4	Teacher and Secr'l. Educ.	215	.42	.66
10	Freshmen	30 - 37	.66	.62
10	Freshmen	239 - 605	.53	.42
11	Science Majors	187	.40	.35

*If two samples were involved, the number on the left is the size of the ACT sample; on the right, the size of the SAT sample. At one school, not all of the N's were available; NA refers to "Not Available."

Appendix 1 (cont.)

The Prediction of Specific College Social Science Course Grades

College	Sample	N	ACT-R	SAT-R	Course
10	Freshmen	99 - 97	.55	.54	Economics
11	Social Science Majors	135	.49	.44	Economics
10	Freshmen	180 - 188	.70	.66	Government
10	Freshmen	130 - 133	.52	.41	History
11	Social Science Majors	107	.46	.32	History
10	Freshmen	NA - NA	.66	.58	Sociology
11	Social Science Majors	130	.38	.42	Sociology
1	Women	139	.44	.42	Social Science
3	Freshmen	125	.31	.34	Social Science
4	Business Majors	261	.23	.23	Social Science

The Prediction of Specific College Science Course Grades

College	Sample	N	ACT-R	SAT-R	Course
1	Men	147	.50	.48	Science
1	Women	254	.46	.46	Science
3	Freshmen	114	.31	.30	Science
10	Freshmen	NA - 76	.55	.54	Anatomy
10	Freshmen	70 - 72	.62	.55	Biology
10	Freshmen	62 - 53	.72	.70	Chemistry
10	Freshmen	201 - 232	.65	.62	Chemistry
11	Science Majors	186	.35	.33	Chemistry
10	Freshmen	63 - 57	.40	.50	Geology
10	Freshmen	NA - NA	.69	.65	Zoology
11	Science Majors	187	.45	.41	Physics

Appendix 1 (cont.)

The Prediction of Other College Grades

College	Sample	N	ACT-R	SAT-R	Course
4	Business Majors	260	.16	.35	Accounting grades
4	Teacher and Secr'l. Educ.	215	.22	.47	Accounting grades
10	Freshmen	[illegible]5 - 94	.48	.69	Education
10	Freshmen	48 - [illegible]8	.69	.56	Fine Arts
10	Freshmen	42 - NA	.54	.38	Philosophy
11	Humanities Majors	165	.32	.34	Philosophy
10	Freshmen	NA - NA	.62	.59	Psychology
10	Freshmen	NA - NA	.46	.19	Music
10	Freshmen	NA - NA	.72	.43	Business
10	Freshmen	NA - NA	.45	.37	Foreign Language
11	Humanities Majors	164	.28	.41	Languages
10	Freshmen	NA - NA	.40	.39	Speech

Appendix 2

Comparison of ACT and SCAT

The Prediction of College English Grades

College	Sample	N	ACT-R	SCAT-R
5	Freshmen	255	.56	.45
6	Freshmen	159	.55	.41
7	Medical Secretarial	108	.57	.52
7	Business	135	.53	.45
8	Freshmen	205	.51	.49
9	Freshmen	131	.72	.74

The Prediction of College Social Studies and Humanities Grades

College	Sample	N	ACT-R	SCAT-R	Area
5	Freshmen	227	.64	.62	Soc. Stud.
6	Freshmen	156	.63	.45	Soc. Stud.
7	Business	135	.55	.44	Soc. Stud.
8	Freshmen	197	.36	.24	Soc. Stud.
5	Freshmen	137	.35	.23	Foreign Language
9	Freshmen	135	.53	.53	Theology

The Prediction of College Natural Science and Mathematics Grades

College	Sample	N	ACT-R	SCAT-R	Area
5	Freshmen	121	.54	.44	Science
7	Medical Secretarial	108	.41	.34	Science
8	Freshmen	197	.36	.24	Science
7	Business	135	.51	.58	Math
8	Freshmen	204	.40	.41	Math

Appendix 2 (cont.)

The Prediction of Over-all College Grade Point Average

College	Sample	N	ACT-R	SCAT-R
5	Freshmen	159	.58	.51
6	Freshmen	164	.67	.48
7	Medical Secretarial	108	.40	.36
7	Business	135	.54	.52
8	Freshmen	205	.54	.53
9	Freshmen	135	.69	.70
12	Freshmen	1857	.50	.49
13	Freshmen	1457	.48	.42
14	Freshmen	263	.49	.44
15	Freshmen	647	.61	.57
16	Freshmen	1875	.51	.47
17	Freshmen	1371	.59	.55
18	Freshmen	66	.53	.44
19	Freshmen	387	.54	.52
20	Freshmen Men	175 - 338	.47	.42
20	Freshmen Women	223 - 455	.54	.45
21	Freshmen	149	.30	.25

ACT RESEARCH REPORT

No. 69

69

August 1975

(insert 30-15pp) (both orders)

PREDICTIVE VALIDITY OF THE ACT TESTS AT SELECTIVE COLLEGES

O. T. Lenning

PUBLISHED BY THE RESEARCH AND DEVELOPMENT DIVISION
THE AMERICAN COLLEGE TESTING PROGRAM
P. O. BOX 168, IOWA CITY, IOWA 52240

PREDICTIVE VALIDITY OF THE ACT TESTS AT SELECTIVE COLLEGES

ABSTRACT

Three studies, each dealing with an aspect of comparative validity of ACT and SAT at selective colleges, are included. The first study considered the predictive efficiency of the ACT test scores and ACT test scores plus high school grades at 120 colleges, separated into three groups according to average college ACT Composite. Predictive efficiency was not found to vary appreciably at various points on the score scale. For 40 colleges where the mean ACT Composite was 24.5 or higher, the median multiple R using the ACT test scores was .46 and using the ACT test scores plus high school grades was .58. By contrast, at the 40 colleges where the mean ACT Composite was 15.5 to 20.0, the median multiple R using the ACT test scores was .46 and using the ACT test scores plus high school grades was .56. The difference is negligible. The median correlation of .58 using ACT test scores and high school grades at selective colleges compares favorably with .54 reported in CEEB materials as the median correlation using SAT test scores and high school rank at colleges defined by the same level of selectivity. Further, when the 40 colleges with the high mean ACT Composites are analyzed moreclosely, there is not a tendency for the predictive efficiency of ACT test scores and high school grades to decline as college mean ACT Composite goes up. The second study concerned four selective colleges where all students had taken both the ACT and the SAT. At all four of these colleges, the mean ACT Composite score was above 24.5; the mean SAT Total score was above 1200. In all four cases, the ACT test scores gave a better prediction of freshman overall Grade Point Average than did SAT. The median R with ACT was .407, with SAT, .316. The third study took place at the U.S. Air Force Academy, a highly selective institution. A focus of concern in this study was to correct for selection on one of the tests, in this case the SAT, by designing the study in such a way that prior selection of students by the SAT would not be an extraneous factor in the analysis of the comparative predictive validity of ACT and SAT. Using the conventional formulas for correction, the comparison of correlation coefficients again favored the ACT over the SAT, .56 to .52 for Sample 1, and .51 to .43 for Sample 2. Finally, when CEEB Achievement tests in English and mathematics were added to the two SAT scores in a multiple, the R obtained was comparableto or below that achieved by the ACT tests alone. The evidence in the three studies points to the conclusion that ACT and SAT scores typically yield similar results at selective colleges, and where they do not the ACT is usually favored with higher correlation coefficients.

PREDICTIVE VALIDITY OF THE ACT TESTS AT SELECTIVE COLLEGES

Oscar T. Lenning[1]

When The American College Testing Program (ACT), with its college admissions and guidance battery, was introduced at the beginning of the last decade, some college educators were concerned about whether this new battery would predict freshman grades as well as older examinations such as the Scholastic Aptitude Tests (SAT). The consensus of studies published since that time, however, has been that the ACT Assessment predicts grades for typical college populations generally as effectively as or better than the SAT battery (Boyce & Paxson, 1965; Burns, 1964; Chase, et al., 1963a, 1963b; Lins, Abell, & Hutchins, 1966; Lenning & Maxey, 1972; Munday, 1965; Passons, 1967; Zimmerman & Michael, 1967). Furthermore, by 1971 more than 2,000 higher education institutions were participating in the ACT Program (The American College Testing Program, 1971, p. 3).

In spite of the general acceptance of the ACT Assessment, one question has not been answered to everyone's satisfaction. Although they would agree that ACT is as efficient a predictor as SAT for typical colleges, it has been the subjective contention of some that ACT should *not* predict as well as SAT for highly selective colleges having a preponderance of students with exceptional academic ability. There has been no objective evidence to support such a belief; rather it has been based on the fact that, unlike the SAT, the ACT was *not* specifically designed for use by highly selective colleges.

It was the purpose of this project to collect all available objective information bearing on the question of comparative predictive validity of the two tests at selective colleges. Three separate studies provide such data, and they are summarized here.

The first study presents correlations typically obtained with ACT data at colleges having quite different ability levels. This not only addresses the question as to whether validity varies appreciably at various points on the score scale, but also permits comparison with validity figures on the SAT reported by Angoff (1971). The second deals with four selective colleges where all students had taken both the ACT and SAT, and comparison of predictive validities was thus possible. The third is a case study of one selective institution, the U.S. Air Force Academy. In this study corrections were made for selection, so that this extraneous factor would not bear on the results.

---

[1] The author gratefully acknowledges several people who helped with this report. First is Risdon J. Westen of the United States Air Force Academy, coauthor of the third study reported here. Second are Nancy S. Cole, Leo A. Munday, and E. James Maxey, all of the ACT Research and Development Division. Formerly a member of the ACT Research and Development staff, the author is now Senior Staff Associate with the National Center for Higher Education Management Systems/Western Interstate Commission on Higher Education.

Study 1: Predictive Efficiency of the ACT Tests at Selective Colleges

Samples and Design

Basic and Standard Research Service[2] records for the years 1970-72 were searched to identify colleges having an ACT Composite score mean of 24.5 or above. This score was chosen as a cut-off because it corresponds to a SAT Total score of 1,100, using the Chase and Barritt Table of Concordance (1966), and permits comparison of ACT data with the SAT data reported by Angoff (1971) for colleges with a SAT Verbal mean of 550 or higher.

Forty colleges having an ACT Composite mean score of 24.5 or above were found and included. For colleges participating more than once in the ACT predictive research services during those 3 years, their latest data were used for the study. The freshman student group sizes ranged from 98 to 4,976, with an average of 740. Because colleges participating in these research services are instructed to include either all or a representative sample of their freshman class, it was assumed that the groups were representative of the entering freshmen at those colleges.

Next, equal size samples of medium and low ability colleges were sought for comparison with the 40 high ability colleges. All colleges participating in the 1972 Basic and Standard Research Services were listed in ACT college code number order. A table of random numbers was used to select a starting point and the total number of colleges was divided by 39 to determine the number of colleges to skip each time before selecting a college. (This procedure in effect provides random selection within geographic strata because each state has a certain range of code numbers assigned to it.) If the college fell into the ACT Composite mean score range of 20.0 to 24.5, it was placed into the medium comparison group. If not, adjacent colleges in the ordered list were checked, alternating front and back, until a college falling into the proper score range was found. Then the proper number of colleges from the point of landing was skipped and another medium college selected in like manner. (Institutions such as nursing schools, business schools, and vocational schools were not included.) This process was carried on until a group of 40 colleges was selected.

The table of random numbers was used once again to select a new initial starting point. Identical procedures were then used to select a stratified random group of 40 colleges with ACT Composite means between 15.5 and 20.0.

Frequency tabulations of validity correlations for the three groups were prepared, and a median correlation was calculated for each group. This was done for ACT test scores as predictors (T-Index multiple correlations), and also for ACT test scores plus high school grades as predictors (TH-Index multiple correlations).

An additional procedure that involved only the high ACT Composite score group of colleges was used. Frequency tabulations of validity correlations were prepared separately for six different ACT Composite score mean levels. Once again, this was done for both T-Index and TH-Index multiple correlations.[3]

Results and Conclusions

The results of the study are presented in Tables 1 and 2. Table 1 presents validity correlation distributions and medians for the three groups of colleges. Considering the restricted range or greater homogeneity for the high ability colleges in comparison to the other two groups, the true predictive efficiency for the high ability group would seem to be comparable to that for the medium ability group and better than that for the lower ability group. Such a finding implies that the predictive validity of the ACT tests is as fully satisfactory for use at selective institutions as it is at more typical colleges and universities.

Angoff (1971) reports median multiple correlations of .52 for men and .56 for women at selective colleges using the two SAT scores plus high school rank as predictors. Based on ACT-SAT equivalency tables, Angoff's colleges and our 40 colleges with

---

[2] The Basic Research Service and the Standard Research Service are two predictive research services offered each year by ACT. College officials can use reports provided by these services to analyze the predictive efficiency of ACT data for their campus, the grading practices on campus, and other factors.

[3] The TH-Index is actually not a multiple correlation. It is developed by obtaining grade predictions based on a multiple using the four tests and on another multiple using the four student-reported high school grades. The grade predictions are then averaged separately for each student. The result for each student is the TH-Index predicted grade. To obtain the TH-Index multiple correlation, the TH-Index predicted grades are correlated against the actual grades received by the students.

TABLE 1

ACT Test Multiple Correlations with Overall GPA for Colleges in Three Separate Ability-Mean Ranges

Validity Correlation Interval	Validity Frequencies					
	ACT Composite Mean 15.5-20.0		*ACT Composite Mean 20.0-24.5*		*ACT Composite Mean 24.5 +*	
	T-Index	TH-Index	T-Index	TH-Index	T-Index	TH-Index
65	1	7	3	17	1	6
60-64	2	6	3	8	0	8
55-59	3	10	5	5	5	15
50-54	5	8	10	3	5	8
45-49	12	0	8	2	14	1
40-44	5	4	5	4	10	2
35-39	5	2	1	1	3	0
30-34	3	2	4	0	1	0
25-29	3	1	1	0	1	0
15-24	1	—	0	—	0	—
Median R T-Index	46		50		45	
Median R TH-Index		56		63		58
Number of Colleges	40	40	40	40	40	40

Note. The top group includes all colleges participating in the ACT predictive research services that had an ACT Composite mean for their students of 24.5 or above. Like numbers of colleges with ACT Composite means in the other two ranges were selected at random as based on their 1971-72 freshmen.

mean ACT Composite scores above 24.5 represent the same level of selectivity. The midpoint for Angoff's correlations is .54, which could be considered as the index of SAT predictive accuracy at selective colleges. In contrast, the median multiple correlation for the ACT TH-Index is .58 for similarly defined selective colleges (see Table 2).

The difference between correlations of .58 and .54 is small. And, the two groups of colleges may not be as comparable as one would think. The ACT-SAT equivalency procedure used in this study to determine what ACT score would correspond to a SAT Verbal of 550, is not precise. It may also very well be that there are some basic differences in characteristics between the ACT selective colleges and the SAT selective colleges which could cause one group to be more predictable on grades than the other group. On the other hand, if such biases do exist, they could just as easily favor SAT as ACT. Therefore, even though the data suggest that ACT prediction compares favorably with SAT prediction at selective colleges (with any difference favoring ACT), we must wait for predictive studies using the two tests on the same students for a definite conclusion.

Table 2 presents validity correlation distributions and medians for specific ACT Composite mean ranges for the high group of colleges. The numbers of colleges in the various cells are too small to draw definitive conclusions, but there *does not seem* to be substantial interaction between predictability and ACT Composite mean score level.

TABLE 2

ACT Test Multiple Correlations with Overall GPA at 40 High Ability Colleges according to ACT Composite Mean Level

Validity Correlation Interval	Total Group Validity Frequencies		ACT Composite Mean 24.5-24.9		ACT Composite Mean 25.0-25.4		ACT Composite Mean 25.5-25.9		ACT Composite Mean 26.0-26.4		ACT Composite Mean 26.5-26.9		ACT Composite Mean 27.0 +	
	T-Index	*TH-Index*	*T-Index*	*TH-Index*	*T-Index*	*TH-Index*	*T-Index*	*TH-Index*	*T-Index*	*TH-Index*	*T-Index*	*TH-Index*	*T-Index*	*TH-Index*
65-69	1	6	1	2	0	2	0	0	0	1	0	1	0	0
60-64	0	8	0	3	0	3	0	0	0	1	0	0	0	1
55-59	5	15	2	6	3	5	0	3	0	0	0	0	0	1
50-54	5	8	2	3	0	2	0	1	1	1	1	1	1	0
45-49	14	1	4	0	6	0	1	0	1	0	1	0	1	1
40-44	10	2	4	0	2	0	2	0	1	2	0	0	1	0
35-39	3	—	1	—	1	—	1	—	0	—	0	—	0	—
30-34	1	—	0	—	0	—	0	—	1	—	0	—	0	—
25-29	1	—	0	—	0	—	0	—	1	—	0	—	0	—
Median R T-Index	46		47		47		42		42		50		47	
Median R TH-Index		58		58		59		46		52		60		70
No. of Colleges	40	40	14	14	12	12	4	4	5	5	2	2	3	3

Study 2: Validity Comparison of ACT and SAT on Same Students at Selective Colleges

Data in a study previously published by Lenning and Maxey (1973) suggest that the ACT battery can predict grades at selective colleges at least as well as the SAT battery. All of the ACT Standard Research Service records for the 3 years from 1969 through 1972 were searched, and 17 colleges that had included SAT scores as Local Predictors in their studies were found. For these institutions, ACT and SAT data were available for the same students. It was found that ACT had decidedly better prediction than SAT at over half of the colleges, but that SAT was a definitely better predictor at only one of the colleges.

Four of the 17 colleges happened to be selective institutions (ACT Composite mean scores above 24.5 and SAT-Total means above 1200); all students in each college had taken both ACT and SAT. As shown by the multiple correlations given in the two right-hand columns of Table 3, in all four cases, ACT gave better prediction of freshman overall GPA than SAT.

Some of the colleges used different tests for selection. Colleges A and B used ACT, College C used SAT, and College D used ACT and SAT (whichever the student took first). This is pertinent because the test used for selection will generally yield a lower correlation with a criterion than another equally good predictor available (Gulliksen, 1950). For this reason the overall results in Table 3 are conservative, because if formulas for prior selection were applied, the disparity between the median ACT and SAT Rs would increase.

TABLE 3

Validity Comparison of ACT and SAT

	N	ACT Composite Mean	SAT Total Mean	R_{ACT}	R_{SAT}
College A	619	27.4	1250	.421	.307
College B	116	26.1	1263	.325	.257
College C	299	25.8	1212	.392	.325
College D	1,159	24.6	1206	.473	.410
Median				.407	.316

The only conclusion one can draw from this evidence is that when the same students at selective colleges are tested with both ACT and SAT, ACT is at least as efficient a predictor as SAT and generally is better. Of course, we must remember that only four selective colleges were studied, and that they may or may *not* be representative of selective colleges in general. However, for these four colleges, the difference favoring ACT was quite large.

Study 3: Validity Comparison of ACT and SAT on Same Students at a Selective Institution with Corrections Made for Selection[1]

Samples and Procedures

The study took place at the U.S. Air Force Academy. Two different samples were used for the study; the second served as a replication sample. SAT had been required of all entering cadets, but some had also taken the ACT Assessment. A search was made of the ACT Assessment Program files to determine which of the freshmen in 1967-68 and 1968-69 had taken both batteries. Student Sample 1 for the study consisted of the 1967-68 Air Force Academy freshmen who had taken both the ACT and SAT (n = 271); Sample 2 included the 1968-69 freshmen who had taken both ACT and SAT (n = 348). Sample 1 had an ACT Composite mean of 27.7 and a SAT Total mean of 1250; Sample 2 had an ACT Composite mean of 27.2 and a SAT Total mean of 1249. As an indication of how academically able these groups were, one should keep in mind that national norms for enrolled freshman men have a Composite mean of 20.4 for ACT (The American College Testing Program, 1971, p. 67) and a Total mean of 949 for SAT (Angoff, 1971, p. 83).

---

[1] This study was originally published as "Prediction at a Highly Selective Institution after Corrections Have Been Made for Selection: ACT versus SAT" by R. J. Westen and O. T. Lenning, *College and University*, 1973, 49, 68-76. With slight revision, it appears here by permission of the journal editor.

End of freshman year overall GPA was the criterion for the study. Pearson product-moment correlation coefficients (r) were calculated between overall GPA and scores on each battery subtest. In addition, stepwise multiple-regression analyses were conducted and multiple correlations (R) computed. For each battery, the computer program initially entered that predictor variable having the maximum zero-order correlation with the criterion. At each succeeding step, the variable was added which produced the greatest reduction in the error or residual sum of squares or, alternatively, which produced the maximum increase in R^2.

Because of the selective entrance requirements, it was assumed that the observed correlations with GPA would not give good estimates of the predictive efficiency of the two batteries. Increased homogeneity results in lower than normal correlations which underestimate the predictive efficiency. This result might be expected for ACT as well as for SAT (even though SAT was used in selection), considering the high correlation typically found between SAT Total and ACT Composite. Therefore, all the predictor correlations with GPA were corrected for homogeneity using the correction formulas for multivariate selection outlined by Gulliksen (1950, pp. 158-166). (Also see Lord and Novick, 1968, pp. 146-148.) As Gulliksen noted (1950, p. 158), the equations for multivariate selection become "almost prohibitively complex" unless matrix algebra is used, so matrix notation will be used in the remainder of this section.

Although selection to the Air Force Academy is also based on a number of other variables (e.g., physical aptitude, athletic activities index, nonathletic activities index), only two explicit-selection variables were of concern for this study: (a) SAT Verbal plus CEEB English Achievement and (b) SAT Quantitative plus CEEB Mathematics Achievement. Three incidental-selection variables were of concern for correcting the SAT correlations with GPA (SAT Verbal, SAT Quantitative, and College Freshman Overall GPA); and five were of concern for correcting the ACT correlations with GPA (ACT English, ACT Mathematics, ACT Social Studies, ACT Natural Sciences, and College Freshman Overall GPA). Incidental-selection variables are those variables of concern for which there is not a specific cutoff score but for which one would expect homogeneity to be affected because of their sizable correlations with the explicit-selection variables.

If X represents the explicit-selection variables and Y represents the incidental-selection variables and if upper case letters refer to the applicant group while lower case letters refer to the selected group of students, the multivariate-selection equation is:

$$C_{YY} = c_{yy} + c'_{yx}c^{-1}_{xx}C_{XX}c^{-1}_{xx}c_{xy} - c'_{yx}c^{-1}_{xx}c_{xy}$$

where:

C_{XX} and c_{xx} are the variance-covariance matrices for the explicit-selection variables, and C_{YY} and c_{yy} are the variance-covariance matrices for the incidental-selection variables.

Standard deviations and intercorrelations for all explicit and incidental variables were calculated. The variances and covariances were, in turn, computed from these and substituted into the above equation, and the equation was solved. This equation was solved separately for each set of incidental variables and for both student samples under study. The corrected correlations between the predictors and the overall GPA were contained in the resulting C_{YY} matrix. These correlations were then squared and multiplied by 100 to give the percentages of criterion (overall GPA) variance accounted for by the predictors.

Results

Intercorrelations, means, and standard deviations for the two study groups are shown in Table 4. Incidental-variable data were available only for 1968-69 applicants. Since it was known that the academy's applicant group varies little on these variables from year to year, these data were used for making calculations for both of the study groups. The zero-order correlation between the two explicit-selection variables for the applicant group (a large part of which was not selected for admission) was .6. The explicit-variable means and standard deviations for the applicant group were as follows:

	Mean	S.D.
SAT Verbal plus CEEB English Achievement	1065.7	165.3
SAT Quantitative plus CEEB Mathematics Achievement	1215.4	169.9

One should note how able and homogeneous the two student groups were, in comparison to the applicant group.

Table 4 shows that three of the four Student Sample 1 ACT subtest correlations with GPA are appreciably larger than both SAT subtest correlations with GPA. For Student Sample 2, all four ACT correlations with GPA are larger than both

TABLE 4

Intercorrelations, Means, and Standard Deviations for the Two Groups of Cadets

(Student Sample 1 values are above the diagonal and Student Sample 2 values are below the diagonal)

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Student Sample 1 Means**	**Student Sample 1 S.D.'s**
1 SAT-V + CEEB Eng. Ach.		34	88	33	51	23	51	36	27	1160.4	114.7
2 SAT-Q + CEEB Math. Ach.	32		29	91	22	50	16	26	35	1341.5	119.7
3 SAT Verbal	87	29		30	41	27	53	43	28	580.0	66.6
4 SAT Quantitative	31	88	28		24	51	17	27	29	670.4	60.3
5 ACT English	52	19	39	20		29	41	35	24	23.9	2.6
6 ACT Mathematics	21	50	19	52	31		35	51	39	30.2	3.0
7 ACT Social Studies	42	06	46	12	42	24		57	33	27.2	3.2
8 ACT Natural Sciences	39	22	43	24	38	33	52		32	28.8	3.3
9 College Overall GPA	16	29	20	22	24	33	28	30		2.7	0.6
Student Sample 2 Means	1166.6	1321.5	587.6	661.3	23.2	29.5	27.0	28.6	2.7		
Student Sample 2 S.D.'s	112.9	110.3	61.4	54.4	2.8	3.0	3.5	3.6	0.6		

Note. The correlations have been rounded to the nearest hundredth and the decimal points deleted.

SAT correlations with GPA. These results would be expected to change after corrections for selection have been made.

Table 5 gives the observed correlations with GPA, the corrected correlations with GPA, and the adjusted percentage of variance accounted for by each predictor. The ACT data accounted for more of the overall GPA variance than did the SAT data for both student samples: 31.8% versus 27.4% for Student Sample 1, and 26.1% versus 18.4% for Student Sample 2. Persons more interested in predictive correlations should note that the corrected multiple Rs for SAT are .523 and .429, while those for ACT are .564 and .511 for Student Sample 1 and Student Sample 2, respectively.

Also of interest to Air Force Academy officials was how much the CEEB English Achievement and Mathematics Achievement Test scores added to the prediction obtained with only the SAT scores. Officials felt that the additional half day of testing for each student should add appreciably to the predictive efficiency of the SAT scores alone.

Table 6 gives observed correlations with GPA, corrected correlations with GPA, and the adjusted percentage of variance accounted for when the CEEB Achievement scores were added as predictors along with the SAT scores. The addition of CEEB Achievement scores for Student Sample 1 brought the corrected multiple R almost up to the corrected R obtained with the ACT tests. For Student Sample 2, however, the corrected SAT correlation after the achievement tests had been added as predictors was still appreciably below that for the ACT tests.

Discussion

The results suggest that ACT scores can be at least as predictive, and likely more predictive, of grades at highly selective institutions than SAT scores. This conclusion seems even more evident when one considers that data at colleges reported in Munday's study (1965) suggested the possibility

91-170 O—82——47

TABLE 5

Observed Correlations with Overall GPA, Corrected Correlations with Overall GPA, and Percentage of the Criterion Variance Accounted for

Predictor Variable	Observed Correlation with GPA (R_o)	Corrected Correlation with GPA (R_c)	% of GPA Variance Accounted for (R_c^2)	Predictor Variable	Observed Correlation with GPA (R_o)	Corrected Correlation with GPA (R_c)	% of GPA Variance Accounted for (R_c^2)
Student Sample 1 (N = 271)[a]							
SAT Verbal	.262	.452	20.4	ACT English	.237	.380	14.4
SAT Quantitative	.288	.469	22.0	ACT Mathematics	.387	.503	25.3
				ACT Social Studies	.326	.441	19.4
				ACT Natural Sciences	.321	.436	19.0
SAT Multiple Regression Analysis	.354	.523	27.4	ACT Multiple Regression Analysis	.444	.564	31.8
Student Sample 2 (N = 348)[a]							
SAT Verbal	.201	.353	12.5	ACT English	.241	.357	12.7
SAT Quantitative	.223	.398	15.8	ACT Mathematics	.327	.444	19.7
				ACT Social Studies	.277	.349	12.2
				ACT Natural Sciences	.302	.401	16.1
SAT Multiple Regression Analysis	.265	.429	18.4	ACT Multiple Regression Analysis	.408	.511	26.1

[a] As indicated in the "Samples and Procedures" section of this study Sample 1 includes all 1967-68 Air Force Academy freshmen who had taken both ACT and SAT for admissions purposes, while Sample 2 includes all 1968-69 Air Force Academy freshmen who had taken both ACT and SAT for admissions purposes

TABLE 6

Observed Correlations with Overall GPA, Corrected Correlations with Overall GPA, and Percentage of the Criterion Variance Accounted for When CEEB Achievement Tests Are Added to SAT

Predictor Variable	Observed Correlation with GPA (R_o)	Corrected Correlation with GPA (R_c)	% of GPA Variance Accounted for (R_c^2)	Predictor Variable	Observed Correlation with GPA (R_o)	Corrected Correlation with GPA (R_c)	% of GPA Variance Accounted for (R_c^2)
Student Sample 1							
CEEB English Ach.	.186	.388	15.1				
CEEB Mathematics Ach.	.346	.507	25.7				
SAT Multiple Regression Analysis	.354	.523	27.4				
SAT + CEEB Ach. Tests Multiple Regression Analysis	.404	.555	30.8	ACT Multiple Regression Analysis	.444	.564	31.8
Student Sample 2							
CEEB English Ach.	.089	.275	7.6				
CEEB Mathematics Ach.	.285	.438	19.2				
SAT Multiple Regression Analysis	.265	.429	18.4				
SAT + CEEB Ach. Tests Multiple Regression Analysis	.285	.466	21.7	ACT Multiple Regression Analysis	.408	.511	26.1

that ACT might more often predict better than SAT for women, and that this study involved only men. It should be remembered, however, that this study pertains only to one institution, an institution with a very specialized purpose. Whether the same results would occur at the more prevalent types of selective colleges and universities must be determined by similar research in those types of institutions.

Another important limitation of this study, one which also appears to be a problem of some of the studies cited, should be mentioned. ACT's experience has been that students who take both the ACT and the SAT tend to have characteristics different from those of students who take only the battery that is required. A sample could be selected at random, and those not submitting scores on the second battery could be tested on this battery on a residual basis so that the sample would be representative of the entire freshman class. Then, however, one would have the problem of a difference in motivation between the testing for the required battery and that for the battery not required for entrance. Testing at different times for different tests (e.g., long periods between the testing and testing occurring at different times of the day) may pose other potential problems. It would seem that an experimental design to overcome such problems is needed.

Conclusions from the Three Studies

Each of the studies reported here contains certain limitations, and applicable limitations have been noted in the discussions of each. However, the bulk of the evidence indicates that ACT and SAT are both valid predictors at selective as well as at more typical colleges. Where ACT and SAT do not yield similar results, ACT is usually favored with higher correlation coefficients. *While it is premature to say that ACT is generally more valid, it is fallacious to state that SAT is generally more valid.*

In conducting any on-campus study comparing ACT and SAT, selective institutions should heed the precautions pointed out in the three studies presented here. It is imperative that student motivation and other testing conditions be controlled for the two test batteries and that corrections be made for selection.

Because both ACT and SAT generally have adequate predictive validity, predictive validity differences between the two batteries perhaps should not be the factor determining which battery will be of most value. Selective institutions (as well as more typical colleges and universities) may judge which of the two testing programs to use on grounds other than relative predictive validity, such as college services provided by the testing agency, usefulness to students, and value in the admissions process.

REFERENCES

The American College Testing Program. *Using ACT on the campus* (1971-72 ed.). Iowa City, Iowa: Author, 1971.

Angoff, W. H. (Ed.). *The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests.* New York: College Entrance Examination Board, 1971.

Boyce, R. W., & Paxson, R. C. The predictive validity of eleven tests at one state college. *Educational and Psychological Measurement,* 1965, **25,** 1143-1147.

Burns, R. L. An investigation of the value of The American College Testing Program, the Scholastic Aptitude Test and the Purdue Placement Tests as predictors of academic success of Purdue University freshmen. *Dissertation Abstracts,* 1964, **25,** 1476-1477.

Chase, C. I., & Barritt, L. S. A table of concordance between ACT and SAT. *Journal of College Student Personnel,* 1966, **7,** 105-108.

Chase, C. I., Ludlow, H. G., Pomeroy, M. C., & Barritt, L. S. *Predicting success for university freshmen.* Indiana Studies in Prediction No. 1. Bloomington: Indiana University, 1963. (a)

Chase, C. I., Ludlow, H. G., Pomeroy, M. C., & Barritt, L. S. *Predicting individual course success for entering freshmen.* Indiana Studies in Prediction No. 2. Bloomington: Indiana University, 1963. (b)

Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.

Lenning, O. T., & Maxey, E. J. ACT versus SAT prediction for present-day colleges and students. *Educational and Psychological Measurement,* 1973, **33**, 397-406.

Lins, L. J., Abell, A. P., & Hutchins, H. C. Relative usefulness in predicting academic success of the ACT, the SAT, and some other variables. *Journal of Experimental Education,* 1966, **35**(2), 1-29.

Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores.* Reading, Mass.: Addison-Wesley, 1968.

Munday, L. A. *Comparative predictive validities of the American College Tests and two other scholastic aptitude tests.* ACT Research Report No. 6. Iowa City, Iowa: The American College Testing Program, 1965.

Passons, W. R. Predictive validities of the ACT, SAT and high school grades for first semester GPA and freshman courses. *Educational and Psychological Measurement,* 1967, **27**, 1143-1144.

Westen, R. J., & Lenning, O. T. Prediction at a highly selective institution after corrections have been made for selection: ACT versus SAT. *College and University,* 1973, **49**, 68-76.

Zimmerman, W. S., & Michael, W. B. A comparison of the criterion-related validities of three college entrance examinations with different content emphases. *Educational and Psychological Measurement,* 1967, **27**, 407-412.

ACT Research Reports

This report is Number 69 in a series published by the Research and Development Division of The American College Testing Program. The first 26 Research Reports have been deposited with the American Documentation Institute, ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D.C. 20540. Photocopies and 35 mm. microfilms are available at cost from ADI; order by ADI Document number. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress. Beginning with Research Report No. 27, the reports have been deposited with the National Auxiliary Publications Service of the American Society for Information Science (NAPS), c/o Microfiche Publications, 440 Park Avenue South, New York, New York 10016. Photocopies and 35 mm. microfilms are available at cost from NAPS. Order by NAPS Document number. Advance payment is required. Printed copies may be obtained from the Publications Division, The American College Testing Program, P.O. Box 168, Iowa City, Iowa 52240, at a cost of $1.00 per copy. Check or money order must accompany request.

The reports in this series published since October 1970 are listed below. A complete list can be obtained from ACT Publications.

No. 37 *Practices and Outcomes of Vocational-Technical Education in Technical and Community Colleges*, by T. G. Gartland, & J. F. Carmody (NAPS No. 01441; photo, $6.80; microfilm, $2.00)

No. 38 *Bayesian Considerations in Educational Information Systems*, by M. R. Novick (NAPS No. 01442; photo, $5.00; microfilm, $2.00)

No. 39 *Interactive Effects of Achievement Orientation and Teaching Style on Academic Achievement*, by G. Domino (NAPS No. 01443; photo, $5.00; microfilm, $2.00)

No. 40 *An Analysis of the Structure of Vocational Interests*, by N. S. Cole, & G. R. Hanson (NAPS No. 01444; photo, $5.00; microfilm, $2.00)

No. 41 *How Do Community College Transfer and Occupational Students Differ?* by E. J. Brue, H. B. Engen, & E. J. Maxey (NAPS No. 01445; photo, $5.50; microfilm, $2.00)

No. 42 *Applications of Bayesian Methods to the Prediction of Educational Performance*, by M. R. Novick, P. H. Jackson, D. T. Thayer, & N. S. Cole (NAPS No. 01544; photo, $5.00; microfilm, $2.00)

No. 43 *Toward More Equitable Distribution of College Student Aid Funds: Problems in Assessing Student Financial Need*, by M. D. Orwig (NAPS No. 01543; photo, $5.00; microfilm, $2.00)

No. 44 *Converting Test Data to Counseling Information*, by D. J. Prediger (NAPS No. 01776; photo, $5.00; microfiche, $2.00)

No. 45 *The Accuracy of Self-Report Information Collected on the ACT Test Battery: High School Grades and Items of Nonacademic Achievement*, by E. J. Maxey, & V. J. Ormsby (NAPS No. 01777; photo, $5.00; microfiche, $2.00)

No. 46 *Correlates of Student Interest in Social Issues*, by R. H. Fenske, & J. F. Carmody (NAPS No. 01778; photo, $5.00; microfiche, $2.00)

No. 47 *The Impact of College on Students' Competence to Function in a Learning Society*, by M. H. Walizer, & R. E. Herriott (NAPS No. 01779; photo, $5.00; microfiche, $2.00)

No. 48 *Enrollment Projection Models for Institutional Planning*, by M. D. Orwig, P. K. Jones, & O. T. Lenning (NAPS No. 01780; photo, $5.00; microfiche, $2.00)

No. 49 *On Measuring the Vocational Interests of Women*, by N. S. Cole (NAPS No. 02071; photo, $5.00; microfiche, $1.50)

No. 50 *Stages in the Development of a Black Identity*, by W. S. Hall, R. Freedle, & W. E. Cross, Jr. (NAPS No. 02072; photo, $5.00; microfiche, $1.50)

No. 51 *Bias in Selection*, by N. S. Cole (NAPS No. 02073; photo, $5.00; microfiche, $1.50)

No. 52 *Changes in Goals, Plans, and Background Characteristics of College-Bound High School Students*, by J. F. Carmody, R. H. Fenske, & C. S. Scott (NAPS No. 02074; photo, $5.75; microfiche, $1.50)

No. 53 *Toward an Integration of Theory and Method for Criterion-Referenced Tests*, by R. K. Hambleton, & M. R. Novick (NAPS No. 02075; photo, $5.00; microfiche, $1.50)

No. 54 *College Student Migration*, by R. H. Fenske, C. S. Scott, & J. F. Carmody (NAPS No. 02215; photo, $5.00; microfiche, $1.50)

No. 55 *Predictions of Performance in Career Education*, by M. R. Novick, P. K. Jones, & N. S. Cole (NAPS No. 02216; photo, $5.00; microfiche, $1.50)

No. 56 *Predictors of Graduation from College*, by E. Nicholson (NAPS No. 02217; photo, $5.00; microfiche, $1.50)

No. 57 *Schooling and Subsequent Success: Influence of Ability, Background, and Formal Education*, by L. C. Solmon (NAPS No. 02218; photo, $5.00; microfiche, $1.50)

No. 58 *Common Fallacies about Heredity, Environment, and Human Behavior*, by A. Anastasi (NAPS No. 02220; photo, $5.00; microfiche, $1.50)

No. 59 *A Study of the College Investment Decision*, by W. W. McMahon, & A. P. Wagner (NAPS No. 02219; photo, $5.00; microfiche, $1.50)

No. 60 *Implementation of a Bayesian System for Decision Analysis in a Program of Individually Prescribed Instruction*, by R. L. Ferguson, & M. R. Novick (NAPS No.: not available at this time.)

No. 61 *Nationwide Study of Student Career Development: Summary of Results*, by D. J. Prediger, J. D. Roth, & R. J. Noeth (NAPS No.: not available at this time.)

No. 62 *Varieties of Accomplishment after College: Perspectives on the Meaning of Academic Talent*, by L. A. Munday, & J. C. Davis (NAPS No.: not available at this time.)

No. 63 *Patterns of Concentration in Large Foundations' Grants to U.S. Colleges and Universities*, by R. Colvard, & A. M. Bennett (NAPS No.: not available at this time.)

No. 64 *Vocational Choice Change Patterns of a National Sample of Community-Junior College Students*, by C. S. Scott, R. H. Fenske, & E. J. Maxey (NAPS No.: not available at this time.)

No. 65 *Considerations and Procedures in National Norming: An Illustration Using the ACT Assessment of Career Development and ACT Career Planning Program, Grades 8-11*, by D. L. Bayless, J. W. Bergsten, L. H. Lewis, & R. J. Noeth (NAPS No.: not available at this time.)

No. 66 *The Measurement of Economic Well-Being in Need Analysis Models*, by W. J. Goggin (NAPS No.: not available at this time.)

No. 67 *Assessing the Career Interests of College Youth: Summary of Research and Applications*, by G. R. Hanson (NAPS No.: not available at this time.)

No. 68 *Sex-Role Socialization and Employment Realities: Implications for Vocational Interest Measures*, by D. J. Prediger, & N. S. Cole (NAPS No.: not available at this time.)

ACT RESEARCH REPORT

No. 80

80

October 1979

Listed 3K-17[illegible] with [illegible]

THE VALIDITY OVER TIME OF COLLEGE FRESHMAN GRADE PREDICTION EQUATIONS

R. Sawyer
E. J. Maxey

PUBLISHED BY THE RESEARCH AND DEVELOPMENT DIVISION
THE AMERICAN COLLEGE TESTING PROGRAM
P. O. BOX 168, IOWA CITY, IOWA 52243

ABSTRACT

This report investigates the validity of prediction equations based on ACT test scores and high school grades. It presents separate results for prediction equations based on ACT test scores only, on high school grades only, and on both kinds of prediction jointly. It also includes a discussion of the relationship between the validity of the prediction equations and the length of time they have been in use.

The study is based on data collected from a random sample of 260 colleges that participated in the ACT Research Services from 1972-73 to 1976-77. Separate prediction equations for each college were calculated from data for the years 1972-73, 1973-74, 1974-75, and 1975-76. These prediction equations were then applied to data from 1976-77 freshmen, and the predicted and actual grades were compared.

The accuracy of the prediction equations based on ACT test scores and high school grades jointly was quite stable over time. The mean absolute error of prediction ranged from .53 to .55 over the four years, the proportion of students whose predicted grade was within .20 of their earned grade ranged from .24 to .25, and the cross-validated correlation ranged from .55 to .56. A similar stability was noted in the accuracy of grade predictions for males and females separately.

Grade predictions based only on ACT test scores had a slightly smaller mean absolute error than grade predictions based only on high school grades, and showed slightly greater stability over time. The mean absolute error of predictions based only on test scores was about .57 for the whole four-year period. For predictions based only on high school grades, the mean absolute error ranged from .57 to .60 over the four-year period.

One can conclude that although most colleges experience some change in their students' academic abilities, their curricula, or their grading practices over a period of time, freshman grade average can in most cases be accurately predicted using equations as old as four years. It is ACT policy that colleges update their prediction equations at least every three years if they are to receive grade predictions for future applicants. The data from this study suggest that more frequent revision is not generally necessary. In considering whether revision of equations is necessary before the required date, college researchers should, of course, determine whether some change has occurred which would necessitate an early revision.

THE VALIDITY OVER TIME OF COLLEGE FRESHMAN GRADE PREDICTION EQUATIONS

Richard Sawyer
E. James Maxey

A problem commonly encountered in predicting college freshman grades from standardized test scores and high school grades is the validity over time of the prediction equations. Changes over time in the distribution of ability among entering students, in a college's entrance requirements, in its freshman curriculum, and in instructors' grading policies can be great enough to make old prediction equations inaccurate. Deterioration of the accuracy of grade predictions has obvious negative implications for both colleges and students. On the other hand, collecting and reporting the data needed to revise prediction equations can be expensive and time-consuming. Therefore, many college researchers would naturally want to use prediction equations as long as possible before updating them.

To accommodate the time schedules of colleges, ACT updates its prediction equations each fall using data collected from students who were freshmen in the previous year. Prediction equations are, therefore, based on freshman grades that are at least one year old. Because most students who take the ACT Assessment in the fall of one year will be college freshmen in the fall of the next year, prediction weights are typically calculated from college grades that are at least two years older than the grades being predicted. To minimize the error from out-of-date equations, ACT requires colleges to participate in its predictive research services at least once every three years if they wish to continue to receive grade predictions for future applicants. Thus, predicted grades are typically based on data that are two to four years older than the grades being predicted.

The primary purpose of this report is to document the relationship between the validity and the age of grade prediction equations based on ACT Assessment test scores and high school grades. A second purpose is to summarize the validity of locally developed grade prediction equations based on ACT test scores only, on high school grades only, and on both kinds of predictors jointly. A third purpose is to suggest various techniques for evaluating local prediction equations.

Prediction equations were calculated for a probability sample of individual colleges using data from the years 1972-73, 1973-74, 1974-75, and 1975-76. The predicted grades from these equations were then compared with the actual freshman grades earned in 1976-77, and the comparisons are stated in terms that are easily interpreted and used.

Previous Research

There is relatively little published research on the rate at which predictive accuracy declines over time. Hills, Klock, and Bush (1965) compared the predicted and earned grades of students at seven colleges in Georgia over a three-year period. They found that the average correlation between the predicted freshman grade average and the actual grade average one and two years later was .64 and .63 respectively. The average correlation between freshman grade average and predictors in the base year was .67.

Bowers and Loeb (1972) found that in predicting grades for freshmen at the University of Illinois, the weights for ACT Composite scores were unstable over a five-year period. They found that the weight for high school percentile rank as a predictor of freshman grades was more stable.

Perrin and Whitney (1976) studied the ACT scores, high school grades, and freshman grades from a national sample of student records. They found very little difference in the accuracy of expectancy tables two and three years older than the freshman grade expectancies they were predicting.

The ACT Assessment Program

The ACT Assessment Program is a comprehensive evaluative, guidance, and placement service for students and educators involved in the transition of students from high school to college. This program is based on the ACT Assessment, which consists of four academic tests, self-reported high school grades, the Student Profile Section (SPS), and the ACT Interest Inventory. The program also includes a number of supporting research services, described below.

The four academic tests of the ACT Assessment measure abilities in the subject areas traditionally identified with college and high school programs: English, mathematics, social studies, and natural sciences. The English Usage Test measures students' understanding and use of the basic elements of correct and effective writing; the Mathematics Usage Test measures their mathematical reasoning and problem-solving ability; the Social Studies Reading Test measures the problem-solving skills required in the social studies; the Natural Science Reading Test measures the critical reasoning and problem-solving skills required in the natural sciences. The arithmetic average of the scores on these four tests is the ACT Composite score, which is often used as a measure of overall academic ability. ACT test scores are reported on a standard scale that ranges from 1 to 36. More detailed descriptive and technical information about ACT test scores can be found in the *Technical Report for the ACT Assessment Program* (1973).

When students register for the ACT Assessment Program, they report the last grade received in each of the four subject areas prior to the senior year of high school. The arithmetic average of these four grades—defined as the high school average—provides another measure of overall academic ability. Maxey and Ormsby (1971) investigated the accuracy of self-reported high school grades and found that about 78% of the students reported their grades correctly.

Another component of the ACT Assessment Program is the Student Profile Section (SPS). Through the SPS, students provide information about their background, extracurricular accomplishments, special academic needs, housing plans, financial need, planned major, and career plans. Students complete the SPS when they register for the ACT Assessment.

The ACT Interest Inventory measures students' preferences for job-related activities in six basic interest dimensions: Science, Creative Arts, Social Service, Business Contact, Business Detail, and Technical. Students complete the Interest Inventory when they register for the ACT Assessment.

Information from all these sources—the ACT tests, high school grades, SPS, and Interest Inventory—is organized into individual Student Profile Reports sent to students and colleges. With the summary information provided by the Student Profile Report, students and educators can make informed decisions and plans.

The ACT Predictive Research Services

ACT offers without charge to colleges two general plans for predicting freshman grades. Each plan is designed to meet the varying needs and resources of colleges which use ACT data.

The Basic Research Service requires minimal effort by a college in reporting data. To participate in the Basic Research Service, college personnel simply mark on a computer-generated roster the overall grade point averages of each of their freshmen. A minimum data base of 100 records is required. Through the Basic Research Service, a college can obtain predictions of overall freshman grade average based on the four ACT test scores and the high school average. If they have data for at least 100 students of each sex, colleges can also obtain separate prediction equations for males and females. During the 1977-78 academic year, 461 colleges reported grades for 126,880 freshmen through the Basic Research Service.

The Standard Research Service is designed for colleges that want to predict specific course grades, develop prediction equations for subgroups of students, or use predictors other than ACT test scores and high school grades. A large variety of research studies can be accommodated by the Standard Research Service. Data can be supplied either through punched cards, optically scanned cards, or magnetic tape. During 1977-78, 185 colleges reported grades for 150,998 students through the Standard Research Service.

Data Base

This study is based on a sample from a data base consisting of student records submitted by institutions through their participation in ACT's predictive research services. The institutions represented in this data base participated in the ACT predictive research services in the academic year 1976-77 and in one or more of the academic years 1972-73, 1973-74, 1974-75, and 1975-76. Therefore, the grades earned by the 1976-77 freshmen at these institutions were available for comparison with grades predicted from equations developed in one or more of the four preceding years. There are 605 colleges represented in the data base.

Because the data in the study were collected from colleges participating in ACT's predictive research services, in some respects they are not representative of students nationally.

- Colleges using the ACT Assessment are located mainly in the Rocky Mountains, Great Plains, South, and Midwest with comparatively fewer in the East/Northeast and on the West Coast.
- Privately controlled institutions are relatively underrepresented among colleges that use the ACT Assessment, and publicly controlled institutions are overrepresented.
- Because participation in ACT's research services is voluntary, the data base is self-selected even among colleges that use the ACT Assessment Program.

The results of the study are, therefore, not necessarily representative of the results that would be obtained if data from all colleges in the nation could somehow be collected. One should be cautious, therefore, in applying the results to institutions which do not use the ACT Assessment or do not participate in ACT's predictive research services. Nevertheless, the study will suggest major trends and extend knowledge in this area beyond the results available to date.

To maximize the number of colleges from which data were available, the variables used in the predictions were restricted to the following specifications: overall grade average as criterion, high school grades and ACT scores as predictors, and subgrouping on sex.

Most of the grade averages in this study are from the first semester of the freshman year. Colleges participating in some of ACT's research services do have the option of pooling grades from previous years or reporting grade averages based on the entire freshman year. ACT does not maintain records of individual colleges' choice of criteria. However, from examining production volumes throughout the year, we estimate that over 60% of the colleges in the data base reported first semester grades for the current academic year, and the rest either reported first year cumulative grade averages or pooled data from previous years.

Sample Design

To reduce the computational costs of this study, weights were calculated and prediction equations were cross-validated on a probability sample of records selected from the above data base. The sampling was carried out in two stages.

First, a random sample of 260 colleges was selected from the 605 colleges in the data base. Weights were computed from all student records submitted by these 260 institutions in the academic years 1972-73, 1973-74, 1974-75, and 1975-76. Observe that not every institution supplied data in every year. The number of colleges by year in the data base and sample is displayed in Table 1.

TABLE 1

Summary of Data Base and Sample for Cross-Validation Study

Year	Number of colleges in data base	Number of colleges in sample	Number of student records for computing weights	Number of 1976-77 student records for cross-validation
1972-73	451	187	97,985	13,732
1973-74	484	203	114,331	15,589
1974-75	494	207	108,118	15,412
1975-76	520	211	109,207	15,384
1976-77	605	260	—	—

For each college in the sample, a random subsample of the 1976-77 student records was then selected. The differences between actual 1976-77 grades and predictions based on 1972-73 equations, 1973-74 equations, and so on, were summarized. Thus a cross-validation was made for all of the equations developed from data in the four years preceding 1976-77.

At the time the study was done, 1976-77 grades were the most recent available for cross-validating prediction equations developed from earlier years' data. Computational costs prevented replicating the study using earlier base years and cross-validation years.

It was anticipated that the validity of grade predictions might vary according to differences in the composition and affiliation of the colleges. By taking note of these differences when designing the sample, it was hoped that greater precision in the inferences might be obtained. In this study, separate samples were selected from strata of colleges determined by their control, type, and size. The stratification variables were:

- The *control* of a college, public or private.
- The *type* of a college, as determined by the highest degree level it offers:

  Type 1: Two but less than four years of work beyond Grade 12—includes junior colleges, technical institutes, normal schools.

  Type 2: Only the bachelor's or first professional degree—includes those institutions offering courses of study leading to the customary Bachelor of Arts or Bachelor of Science degree and all those degrees which entitle the possessor to enter the profession indicated.

  Type 3: Master's and/or second professional degree—includes those institutions offering the customary first graduate degree and any degree earned in the same field after the bachelor's or first professional degree. This type of institution does *not* offer the Doctor of Philosophy or equivalent degrees.

  Type 4: Doctor of Philosophy and equivalent degrees—includes those institutions which are considered universities.
- The *size* stratum for a college, as determined by the number of students for which the college reported 1976-77 freshman grades.

Category 1: 100 students or fewer

Category 2: 101-200 students

Category 3: 201-500 students

Category 4: 501-1000 students

Category 5: 1001+ students

Colleges do have the option of reporting grades for a sample of their freshman class, provided the sample size is 100 or larger. Moreover, colleges can pool data from previous years in order to attain the required minimum sample size. Therefore, the size measure in this study is not necessarily the same as total freshman class size. However, for about 70% of the institutions, the number of records in the data base is within 30% of the estimated freshman enrollment.

The subsampling rates for 1976-77 records were a function of the size strata for the college sample. These subsampling rates and the number of colleges selected from each size stratum are displayed in Table 2.

These sample sizes were chosen to yield a 95% chance that the mean absolute error (defined below) estimated from the sample would be within the following limits of the mean absolute error computed from all records in the data base:

- ± .01 grade units of the mean absolute error of all students in all colleges.
- ± .03 grade units of the mean absolute error for all students in each size stratum.

The second condition was imposed to permit the use of the sample to study the effects of college size on predictive accuracy (results not reported here). Examination of the sampling variances estimated from the data indicated that these precisions were attained.

TABLE 2

Stratification and Subsampling for the Cross-Validation Study Sample by Size Stratum

Size stratum	**Number of colleges in data base**	**Number of colleges in sample**	**Subsampling rate for 1976-77 records**
100 or fewer	47	38	1/2
101-200	221	75	1/4
201-500	202	63	1/4
501-1000	70	44	1/6
Over 1000	65	40	1/16
Total	605	260	–

Prediction Equations

Prediction equations were calculated from the 1972-73, 1973-74, 1974-75, and 1975-76 data using a standard four-variable multiple linear regression of college freshman grade average on the four ACT test scores:

$y = a_0$

$+a_1 \times$ ACT English score

$+a_2 \times$ ACT Mathematics score

$+a_3 \times$ ACT Social Studies score

$+a_4 \times$ ACT Natural Science score

where a_0, a_1, a_2, a_3, and a_4 are regression weights calculated from the base year data. In the ACT Standard Research Service this prediction is called the T Index.

Prediction equations were also calculated using:

- the four self-reported high school grades only in a multiple linear regression equation (H Index)
- the average of the T Index and H Index (the TH Index)
- the four ACT test scores and the four high school grades in an eight-variable multiple linear regression equation.

Past research has shown that the correlation between the TH Index and freshman college grade average is only slightly less than that for the traditional eight-variable multiple linear regression equation (*Technical Report for the ACT Assessment Program*, 1973).

Previous research also suggests that the relationship between high school grades, standardized test scores, and freshman grade average differs for males and females (*Technical Report for the ACT Assessment Program*, 1973). Moreover, colleges often elect to receive separate predictions by sex. For these reasons, prediction equations were calculated separately for males and females in each college, as well as for all students in a college.

Some colleges in the sample submitted student grade data without identifying the sex of their students. Therefore, the sum of the number of males and females from which the separate-sex equations were computed is less than the total number of records from which the combined-sex equations were computed.

Cross-Validation Statistics

The difference between the predicted and actual 1976-77 grade average for a student is called the error of prediction. The magnitude of this error, ignoring the direction, is called the absolute error of prediction. If the absolute error of prediction is averaged over a group of students (for example, all the students in a college), the resulting number is called the mean absolute error (MAE) of prediction in that group of students. This statistic has an immediate relevance to the quality of grade predictions. For example, if the MAE in predicting grade average in a college is .45, then, on the average, there is an absolute discrepancy of .45 grade units between the predicted and actual grade averages in the college.

A related measure of prediction quality that is sometimes used is the mean squared error, which is the average over a group of students of the squared error of prediction. It is actually this quantity that the usual least-squares regression methods seek to minimize. Because the mean squared error is less directly interpretable in practical terms than the MAE, it was not used here.

Another useful measure of the quality of prediction for a group of students is the proportion whose predicted grade is within a certain range of the actual grade. In this study we report the proportion of students whose predicted grade average is within 0.20 grade units of actual grade average; this proportion will be denoted by the abbreviation P20. An absolute error of 0.20 grade units represents a high degree of accuracy in prediction. The statistic P20, therefore, measures the proportion of students for whom extremely accurate grade predictions were possible. Some researchers may consider prediction errors larger than 0.20 to be quite acceptable, analogous proportions (such as the proportion of students whose predicted grade average is within 0.50 grade units of the actual grade point average) could be computed for them.

Probably the most commonly used measure of predictive efficiency is cross-validated r, that is, the Pearson correlation between the predicted and actual grades. Generally, this coefficient is compared with the correlation coefficient calculated from the base year data to give an indication of the accuracy and stability of the prediction equations. Because it is so widely used, this statistic is reported in addition to MAE and P20.

A limitation of this statistic is illustrated by the hypothetical data in Figures 1a and 1b. These plots suggest that false conclusions could result from using cross-validated r as the sole measure of prediction quality. In both cases, cross-validated r is near 1, yet in both cases the prediction is not very good. In the first instance, the predicted grades are consistently too low; in the second, they are systematically too low for low-ranking students and too high for high-ranking students. Thus, a large cross-validated r does not guarantee success in prediction. A low value of cross-validated r will indicate poor prediction, however, because if the predicted and actual grades have little linear relationship, they cannot be very close to each other.

A justification sometimes advanced for the correlation coefficient is that r^2 measures the proportional decrease in squared error from that which would result if the mean grade were predicted for each student. Similarly, r^2 takes into account the effect on prediction due to the inherent variability of students' grades.

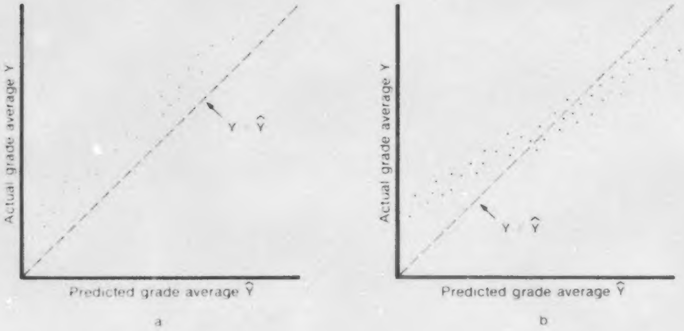


Figure 1. Two situations in which cross-validated r gives a false indication of prediction quality

Computations

The prediction equations and cross-validation statistics for each college were computed by a custom program using the IMSL subroutine package (IMSL, 1977). The program was tested on data from three colleges over a three-year period. Every computation on these data was compared with one obtained from the SPSS statistical package (Nie, Hall, Jenkins, Steinbrenner, and Brent, 1975) and, where feasible, with one computed by hand. In all cases, the cross-validation statistics for each college were within ±.001 of each other, and in a majority of cases they were within ±.0001.

The cross-validation statistics from each college were summarized through SPSS routines. The data were weighted to take into account the differential sampling rates in the various strata.

Results

The results of the analyses are displayed in Tables 3a-3d, 4a-4d, 5, and 6. They indicate that for *students* as a whole (irrespective of their college), the accuracy of all four kinds of prediction equations was quite stable over the four-year period studied. The same is generally true of the accuracy of the prediction equations for *colleges*, although a few individual colleges do show some instability.

Tables 3a-3d display the cross-validation statistics for all students in the sample. Over the four year period, there is no marked increase or decrease in any of the three statistics for the four kinds of predictions. For the T Index, MAE for the total group of students is about .57, P20 is .22-.23, and r is .48-.49. For the H Index, there is slightly more deterioration over time: MAE ranges from .57-.60, P20 from .21-.23, and r from .49-.51. The TH Index predictions have an MAE of .54-.55, a P20 of .23-.24, and an r of .55-.56. The eight-variable regression equations showed virtually the same results: MAE from .53-.55, P20 from .24-.25, and r from .55-.56.

TABLE 3a

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(T Index—ACT Test Scores Only)

Group	**Cross-validation statistic**	**Year in which prediction equation was developed**			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.57	.57	.57	.57
	P20	.23	.22	.23	.23
	r	.48	.48	.49	.49
Males	MAE	.61	.61	.60	.61
	P20	.21	.20	.21	.21
	r	.43	.43	.43	.43
Females	MAE	.56	.55	.55	.55
	P20	.22	.24	.25	.23
	r	.53	.50	.51	.51

TABLE 3b

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(H Index—High School Grades Only)

Group	**Cross-validation statistic**	**Year in which prediction equation was developed**			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.60	.59	.57	.57
	P20	.21	.23	.22	.23
	r	.49	.49	.50	.51
Males	MAE	.62	.61	.60	.60
	P20	.21	.20	.22	.21
	r	.46	.46	.46	.48
Females	MAE	.57	.55	.54	.54
	P20	.24	.23	.24	.24
	r	.50	.53	.54	.53

TABLE 3c

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(TH Index)

Group	**Cross-validation statistic**	**Year in which prediction equation was developed**			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.55	.55	.54	.54
	P20	.23	.23	.24	.24
	r	.55	.55	.56	.56
Males	MAE	.59	.59	.57	.58
	P20	.22	.22	.23	.23
	r	.50	.50	.50	.51
Females	MAE	.53	.52	.52	.52
	P20	.24	.24	.26	.25
	r	.58	.58	.59	.59

TABLE 3d

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(Eight-Variable Multiple Linear Regression)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	55	54	53	54
	P20	24	24	25	24
	r	55	55	56	56
Males	MAE	59	59	58	58
	P20	22	23	23	23
	r	50	50	50	51
Females	MAE	53	52	51	51
	P20	24	25	26	25
	r	57	57	59	59

The T Index predictions were somewhat more stable over time than the H Index predictions. The reasons for this cannot be easily determined, but could be related to changes in high school curriculum or inflation in high school grades.

Although the H Index r's are slightly but constantly larger than the T Index r's, the statistics MAE and P20 indicate that, on the average, the T Index resulted in slightly better predictions than did the H Index. That the H Index r is larger than the T Index r is consistent with earlier published research (*ACT Technical Report*, 1973). The results for MAE and P20, however, suggest that something like the situation illustrated in Figures 1a and 1b occurred in many of the colleges.

Further examination of the prediction errors revealed that the H Index predictions based on the 1972-73, 1973-74, 1974-75, and 1975-76 data underestimated the criterion on the average by 14, 10, 04 and less than 01, respectively. The T Index predictions based on data from these four years often overestimated the criterion, but by an average of only 05, 05, 05, and 03, respectively. Thus, the H Index tended to underestimate the criterion, and the amount of underestimation increased with the age of the prediction equation. The T Index tended to overestimate the criterion, but by a fairly constant and small amount over time.

Prediction equations developed for males and females separately show the same stability in quality as the total group equations. The predictions for females were somewhat more accurate (eight-variable multiple regression MAE = 51- 53, P20 = 24- 26, r = 57- 59) than the predictions for males (eight-variable multiple regression MAE = 58- 59, P20 = 22- 23, r = 50- 51). This result is consistent with earlier analyses of ACT grade predictions (*ACT Technical Report*, 1973).

It should be noted that these results pertain to separate-sex prediction equations, rather than to a total group equation. The cross-validation results for predictions made from combined-sex equations, however, are virtually identical to those for predictions made from the separate-sex equations. This would indicate that there is little average benefit in developing separate-sex equations.

Tables 4a-4d display the college medians of the cross-validation statistics. The numbers in this table show that measures of prediction quality for students across all colleges are also typical of the measures of prediction quality for individual colleges. The same stability in prediction quality over the four years is evident.

TABLE 4a

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(T Index—ACT Test Scores Only)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	55	56	55	55
	P20	23	23	22	22
	r	48	48	48	48
Males	MAE	60	61	60	60
	P20	22	24	20	20
	r	44	42	44	42
Females	MAE	54	54	53	54
	P20	23	24	24	23
	r	50	52	53	51

TABLE 4b

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(H Index—High School Grades Only)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	58	57	60	54
	P20	21	21	23	23
	r	48	50	49	50
Males	MAE	60	62	59	59
	P20	20	20	21	22
	r	45	47	45	48
Females	MAE	57	54	53	53
	P20	21	22	24	23
	r	54	53	54	55

TABLE 4c

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(TH Index)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.53	.54	.52	.52
	P20	.24	.23	.24	.24
	r	.55	.56	.56	.56
Males	MAE	.56	.59	.56	.56
	P20	.22	.21	.23	.23
	r	.54	.54	.53	.55
Females	MAE	.52	.51	.50	.50
	P20	.24	.24	.25	.24
	r	.58	.60	.60	.59

TABLE 4d

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(Eight-Variable Multiple Linear Regression)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.53	.53	.52	.51
	P20	.24	.24	.25	.25
	r	.54	.55	.56	.56
Males	MAE	.57	.60	.57	.56
	P20	.21	.21	.21	.23
	r	.52	.51	.51	.51
Females	MAE	.52	.51	.50	.52
	P20	.25	.25	.26	.25
	r	.57	.57	.59	.58

Table 5 presents a frequency distribution of MAE for the eight-variable regression among colleges for the four years studied. For about 72%-78% of the colleges, the MAE is 0.60 or less. Thus, the MAE reported in Table 3d for students over all colleges is fairly typical of the MAEs in most colleges individually. For a small proportion (2%-4%) of colleges, MAEs in the range 0.80-1.10 occurred.

The relative stability of the quality of predictions over all students and colleges could mask instability in particular colleges. Table 6, which addresses this issue, is a frequency distribution of the differences between MAE for eight-variable multiple regression equations developed in 1975-76 and MAE for the equations developed in 1974-75, 1973-74, and 1972-73. The results indicate that for most colleges, MAE is quite stable from year to year. The number of colleges with a larger MAE from 1974-75 prediction equations than from 1975-76 prediction equations is approximately the same as the number of colleges with a smaller MAE from 1974-75 prediction equations than from 1975-76 equations. Moreover, about 99% of the differences in MAE are in the range -.10 to +.10. Slightly more colleges have a larger MAE from equations developed in 1974-75, 1973-74, or 1972-73 than 1975-76, but the differences in MAE are concentrated in the range of -.10 to .00. In summary, the MAE for over 90% of the colleges remains stable, even over a period of four years.

TABLE 5

Proportion of Colleges with Various Ranges of Mean Absolute Error in Predicting 1976-77 Freshman Grade Point Average

(Eight-Variable Multiple Linear Regression)

Range in MAE	**Year in which prediction equation was developed**			
	1972-73	1973-74	1974-75	1975-76
0.0 - 0.1	00	00	00	00
0.1 - 0.2	00	00	00	00
0.2 - 0.3	00	00	00	00
0.3 - 0.4	07	08	12	09
0.4 - 0.5	31	31	29	36
0.5 - 0.6	34	34	37	33
0.6 - 0.7	18	21	15	13
0.7 - 0.8	09	05	06	05
0.8 - 0.9	02	02	01	02
0.9 - 1.0	00	00	00	01
1.0 - 1.1	00	00	00	01

TABLE 6

Stability of Mean Absolute Error across Four Years: Proportions of Colleges with Various Differences in MAE from 1975-76 Equations and MAE from Equations Developed in Earlier Years

(Eight-Variable Multiple Linear Regression)

Range of difference in MAE from 1975-76 equations and MAE from older equation	Earlier year in which prediction was developed		
	1972-73	1973-74	1974-75
-0.50 to -0.40	.00	.00	.00
-0.40 to -0.30	.00	.00	.00
-0.30 to -0.20	.00	.00	.00
-0.20 to -0.10	.06	.03	.01
-0.10 to 0.00	.63	.69	.55
0.00 to 0.10	.30	.28	.44
0.10 to 0.20	.00	.00	.00
0.20 to 0.30	.01	.00	.01
0.30 to 0.40	.00	.01	.00
0.40 to 0.50	.00	.00	.00
0.50 to 0.60	.00	.00	.00

Summary and Conclusions

The stability of prediction equations based on ACT scores and high school grades was investigated for a stratified random sample of 260 colleges that participated in the ACT Research Services during the period of 1972-73 to 1976-77. Separate prediction equations for each college were calculated from data for the years 1972-73, 1973-74, 1974-75, and 1975-76 and applied to 1976-77 freshmen. Prediction errors for 1976-77 grade averages were then summarized.

The results indicate that for most students and colleges, the quality of the prediction equations is quite stable over the four years. The mean absolute error of prediction using an eight-variable multiple regression equation ranged from .53 to .55, the proportion of students whose predicted grade was within .20 of their earned grade ranged from .24 to .25, and the cross-validated correlation ranged from .55 to .56. A similar stability was noted in the accuracy of predictions for males and females using separate-sex prediction equations.

Grade predictions based on test scores only had a slightly smaller mean absolute error than grade predictions based on high school grades only and showed slightly greater stability over time. The mean absolute error of predictions based on test scores only was about .57 for the whole four year period. For predictions based on high school grades only, it ranged from .57 to .60 over the four year period.

One can conclude from this study that although most colleges experience some change in their students' academic abilities, their curricula, or grading practices over a period of time, freshman grade average can in most cases be accurately predicted using equations which have been in use for as long as four years. It is ACT's policy that colleges update their prediction equations at least every three years if they are to receive grade predictions for future applicants. The data from this study suggest that more frequent revision is not generally necessary. In considering whether revision of equations is necessary before the required date, college researchers should, of course, determine whether some change has occurred which would necessitate an early revision.

REFERENCES

The American College Testing Program. *Technical report for the ACT Assessment Program.* Iowa City, Iowa: Author, 1973.

Bowers, J., & Loeb, J. Changes in predictor weights in GPA regression equations. *Measurement and Evaluation in Guidance,* 1972, 5, 438-442.

Hills, J., Klock, J., & Bush, M. The use of academic prediction equations with subsequent classes. *American Educational Research Journal,* 1965, 2, 203-206.

International Mathematical and Statistical Laboratory. *IMSL library 1 reference manual* (6th ed.). Houston, Texas: Author, 1977.

Maxey, E. J., & Ormsby, V. J. *The accuracy of self-report information collected on the ACT test battery: High school grades and items of nonacademic achievement* (ACT Research Report No. 45). Iowa City, Iowa: The American College Testing Program, 1971.

Nie, N., Hull, C. H., Jenkins, J., Steinbrenner, K., & Bent, D. *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.

Perrin, D. W., & Whitney, D. R. Methods for smoothing expectancy tables applied to the prediction of success in college. *Journal of Educational Measurement,* 1976, *13,* 223-231.

Mr. **Weiss**. What kind of request form do you have for the test takers if they want to get copies of the answer sheet or results or the test questions?

Dr. **Rever**. In New York, we provide a special insert to our examination and students are informed about the procedures for obtaining copies of the examination. We have recently changed the procedures so that there is an order blank on that special insert that can be completed and returned to Iowa City. We would anticipate as a result of that, indeed, we will receive increased percentage of students who request copies of the ACT assessment. However, at the very, very low rate in which we have received rates, that increase would have to be magnified many, many times over to get a rate of request for materials comparable to that of the college boards or the one that its——

Mr. **Weiss**. Can that order form be simply a request form—be included with the test when the test taker completes it or does it have to be separately taken and mailed once the test taker leaves the examination room?

Dr. **Rever**. I believe the question you're asking is, is that a part of the standard registration document that students submit to us to register for the examination. No, it is not. It is a separate document that would be submitted separately.

Mr. **Weiss**. Or is it a document that in fact is part of the examination packet itself, that when——

Dr. **Rever**. Right.

Mr. **Weiss**. A person takes the examination can simply fill out the form and send—and hand it in with the test that he or she is taking?

Dr. **Rever**. That's possible though under our current procedures—well, all I'm saying, we do not do that. In other words, when a student sits for the examination, we do not provide them the opportunity to order copies of the test materials. They are provided that opportunity when they receive their registration materials or they can do it independent of that, whichever they prefer.

Mr. **Weiss**. You were here, I believe, during the testimony of Mr. Anrig and heard, or did you, his challenge and proposal urging all profit and nonprofit testing corporations and those agencies governing which test to join with ETS voluntarily in an industrywide code of fair testing for educational testing nationwide, and there are some other elements? What's your response to that invitation?

Dr. **Rever**. Indeed, Mr. Chairman, we will review that invitation and that challenge with great interest. We will respond to that in due course as we've had an opportunity to review it. I think you and the concerns you've expressed in these hearings are cognizant of the fear of test agencies, not ACT but of other, of being dominated by the values and goals of ETS. ACT has maintained a degree of separateness from that as part of our competitive environment. I think we would be somewhat skeptical in looking at that though the spirit in which it's offered, I believe, is one in which we would find very comfortable and would in fact reward or we find a great deal of comfort in the work of the three associations Mr. Anrig identified in establishing standards for tests. We find that, in the past, they've been conscientious, that the advantage to that is they are third parties to the process of testing. They are not test developers and they are not students, but they are scholars and psychometricians, and we find that may be the best arena or process by which such standards and guidelines may be developed. I don't think Mr. Anrig intended to say that ETS will dominate this process. I think he's cognizant of concerns of other test agencies and their relationship with ETS. So I can promise you that we will earnestly study that, that we will work with Mr. Anrig to bring that to fruition if that seems to be in students' interests and in the interests of education and ACT.

Mr. **Weiss**. And finally, I am correct in listening to your testimony that you make the test questions, answers, and scoring sheets available only to students who take the test in the State of New York. Is that correct?

Dr. **Rever**. That's not correct. If I'll restate. Maybe it is correct but I just misunderstood your question. At the current time, students can obtain hand scoring services at virtually any location throughout this Nation. Students in New York are the only students that can obtain copies of the ACT assessment shortly after it's administered. Students throughout the Nation can obtain retired forms of the ACT assessment, forms that were retired last year. We intend to follow this pattern as we observe developments and our judgment so indicates that that pattern may well be expanded and in fact we continue to look at that with great deal of interest.

Our concern has been, quite frankly, that the very low interest that we've experienced among New York students probably is an accurate estimate of what we would experience nationwide. One must understand, and we have repeatedly emphasized this, our examination is quite different from other examinations. The kinds of students that we test are somewhat different from others—from other test agencies. The use to which the instrument is placed is quite different from other instruments. Not saying that the uses of other instruments are bad. In fact, they are indeed appropriate. It's just that our instrument is used in different ways. We believe students see that. We believe students understand that, and the low request rate which is related to the very low rate at which we find students taking the examination on more than one occasion suggests that we would be better off, for example, by letting students see a copy of the examination before they, you know, the likelihood is they're not going to take the examination more than once. So we have chosen to provide access to examinations before they take it and made that freely available so that if there are benefits to be derived, they would be derived before the student sits for examination.

Mr. **Weiss**. Not the examination that they're going to be taking?

Dr. **Rever**. No, no. This would be a retired form.

Mr. **Weiss**. Right. ETS does the same thing, and has been doing so for years.

Dr. **Rever**. Indeed. Indeed. Right.

Mr. **Weiss**. Do you know if there are coaching courses that are directed toward the ACT tests?

Dr. **Rever**. I've seen and I don't mean to pick on Mr. Kaplan, I've seen an advertisement of Mr. Kaplan's recently that indicated that they are now prepared to help students prepare for the ACT assessment. I recently identified an organization, I believe, in the State of Connecticut that does a mail order home study course that they claim will help students prepare for the ACT assessment. There are some but I cannot really say the number of them. Our guess is they are relatively few.

Mr. **Weiss**. Thank you very much. Mr. Williams?

Mr. **Williams**. Mr. Rever, my concern—our concern, I think, on the committee is the same as yours, and that is No. 1, the tests be useful, and No. 2, that their application not create a negative in education. In trying to make that determination, one tends to rely some on studies that are done of tests and their ramification, I guess, those who—those of us who have some doubt about tests rely on the tests of tests to determine whether or not they're—they're valuable. Let me—let me cite to you one study which was conducted by the University of Missouri's Columbia School of Medicine concerning admissions.

Quoting now from the study, "Neither of the cognitive measures, college GPA's or MCAT's, correlated significantly with letter ratings after 4 years of medical school."

Another study from the University of California at Berkeley stated, quoting, "It must be concluded that ratings of internship performance for those students were not related to and could not be predicted by MCAT scores," and an additional study which was conducted by the American Medical Colleges Association said, quote: "An admissions policy based exclusively on GPA's and/or MCAT scores would proportionately exclude minority applicants from medical education."

I'm quoting now from a study that was done some time ago, in—that last quote, some time ago in 1977. I think test have been improved dramatically since then, and I want to—I want to emphasize that fact. I think tests have been improved dramatically. but I place those quotes in the record because my interest in the chairman's bill goes to (a) the usefulness of the test, (b) their demonstrated relationship to the student's eventual performance as a physician or a dentist or a lawyer, and finally, that we be absolutely certain in this society through whatever means we have to take that those tests for admissions to America's colleges, which means admission to a piece of American pie, not exclude minorities. You may want to comment on my——

Dr. **Rever**. Oh, indeed, and I may have trouble recalling all of the points you made, but let me try and address myself to some very specific ones and you may want to call others to my attention.

Mr. **Williams**. Sure.

Dr. **Rever**. The issue of the usefulness of examinations cannot be determined by correlation coefficients. We have and we encourage, in fact, institutions to do everything within their power to destroy the predictive validity of the ACT assessment. Now you may find that strange. The reason we encourage institutions to do that is because they have, by destroying the predictive validity of the ACT assessment, perhaps used examinations appropriately in helping to guide students or design their curriculum program and provides whatever services they may need in improving their study skills or improving their study development or whatever it may be so that they can beat the predictions.

The object of a prediction is not to guarantee what the future of the student's going to look like it is to indicate to the student and to the institution what kinds of assistance or difficulties the students may indeed encounter.

Consequently, zero correlation coefficients are often more greatly valued by ACT and other test organizations than correlation coefficients in a range of .70 to 1.0 because depending upon the institutions purposes, correlation coefficient that high—predictive validity that high—may be indicating to the institution they are doing a disservice to students and not fulfilling their societal obligations to educating students that deserve to be educated, and with regard to test bias and the selective admissions selections of—the admission of minorities or let's just say the disadvantaged, I would like to call the subcommittee's attention to a recent, and this came across my desk yesterday, issue of the American Psychologist. This is my personal copy. I refuse to give it to you, and in it let me—let me give you the title. Just a minute. It is Testing: Concepts, Policy, Practice and Research.

In this document you will find a contribution by one of the foremost authority on bias and selection and indeed we are talking about bias in selection when we're talking about excluding the disadvantaged from selective educational programs such as law schools and medical schools.

Nancy Cole (phonetic), the University of Pittsburgh, one of the leading observers and scholars in the field of the theory of selection and in psychometrics concluded that indeed no matter what kinds of theoretical or mathematical scheme one employs in selecting students to introduce fairness into the test process that ultimately it is a value judgment that will dictate the results one achieves. In other words, there are no answers to the question of are tests simply biased and used against disadvantaged students. Tests are not biased according to some definitions. According to the most simple-minded they are, but that's a hardly useful definition. What we've observed, I think, in education, particularly in selective institutions over the past few years, is that indeed minorities and indeed the disadvantaged have been represented to a greater degree in entering classes than they were years before. Tests have remained the same. Their performance on tests has been roughly the same although minority students' test scores have been tending to go up certainly not enough to account for the increased proportion of their representation in entering classes among selected institutions.

I, as an observor and some time follower of these phenomena, have not identified any conclusive empirical evidence that would suggest that tests have kept minority and disadvantaged students out of institutions, and in fact the evidence is in the opposite direction. We know, for example, that using standard prediction techniques, it works to the advantage of minorities and disadvantaged to have taken an examination as part of the admissions process because we tend to overpredict their performance in college.

That has many implications and I don't mean to try to make this into an admission psychometrics decision theory statement, but I think one needs to be very careful about the kinds of statements that are made about how tests have affected students when, in fact, there is a wealth of information saying that tests have been helpful, they continue to be helpful, and our intent, and we share this intent with Mr. Williams to make sure they are useful.

I think the evidence to date has tended to support our view that indeed they have been useful and indeed are helpful and valid for their purposes.

Mr. **WILLIAMS**. What—what efforts in the preparation of tests is made by testing community to assure that tests are not biased against significant groups of American citizens?

Dr. **REVER**. That's a very good question, Mr. Williams. I think test agencies, in my short experience—I've been with the American college testing program for 12 years and was in their research and development division for 7 before moving to Washington, D.C. During my experience in our research and development division which develops the ACT assessment and other tests, we go to great pains to see to it that certain kinds of bias are excluded from the examination. We have every test item reviewed by a panel of representatives of minorities and other disadvantaged students to identify and to ferret out objectionable or what might be seen on its face as culturally biased language and concepts and the like. In addition to that, when the examination items are submitted or taken by a representative sample of students for standardization purposes, we identify students according to their race and look at the pattern of responses that they provide to particular items and to the entire examination and compare that patterns of responses to the majority, white students. By doing so, we intend to identify emerging patterns that may be entirely unrelated to the educational background of students, but again may give us some clues to items that may disadvantage particular kinds of students for reasons beyond their control.

There are other kinds of bias that are present but in essense what we've done is we've tried to eliminate what we call facial bias or item bias by having the examination items extensively reviewed by panels of persons and by looking at the statistical characteristics of the item according to the race of students who've taken it under experimental conditions.

Mr. **WILLIAMS**. Thank you. Thank you, Mr. Chairman.

Mr. **WEISS**. Dr. Rever, thank you very much for your testimony and for your patience in getting through the morning with us. You will forward some of those materials that we discussed?

Dr. **REVER**. Indeed I will.

Mr. **WEISS**. If you'll do it within 10 days, it will be appreciated.

Dr. **REVER**. Will do it tomorrow.

Mr. **WEISS**. Good. Thank you so much.

Our next witness will be Mary Ann Austen, legal counsel, representating State Senator Kenneth Lavalle of New York, who I might note belongs to the other party.

Also, the last time you appeared before was just about 1 year ago.

Mrs. **AUSTEN**. Yes.

Mr. **WEISS**. I guess we introduced you as Miss Mary Ann McLean. I'll extend the committee's congratulations and best wishes to you.

Mrs. **AUSTEN**. Thank you very much.

STATEMENT OF MARY ANN AUSTEN, COUNSEL TO STATE SENATOR KENNETH P. LAVALLE OF NEW YORK

Mrs. **Austen**. Mr. Chairman, my name is Mary Ann Austen, and I serve as counsel to Senator Kenneth P. Lavalle, the sponsor of New York's truth-in-lending law.

Let me extend his apologies for not appearing before this committee today; however, he had previously scheduled legislative hearings which could not be rearranged.

I appreciate the opportunity to be here in his place, and I hope to provide you with information which should assist in your decision to report H.R. 1662 favorable out of committee.

It has been almost 2½ years since truth-in-testing legislation was signed into law in New York. The dire predictions of the opposition have not become a reality. The tests are still offered, the students still take them, and the colleges still use them. All of the players, that is the test agencies, the colleges and most importantly the students have accommodated the changes and have gone about their business. By all reports, the atmosphere is quite calm. It is important to emphasize this fact to you, expecially to those of you who remember the early hearings back in the summer of 1979 when the testimony threatened certain chaos from the demise of a time honored system. You might say that we who have always supported test disclosure had more faith in the resilience of the testing system than those who ran it. We believed the process could sustain the necessary changes and we, as you now know, were right.

Our first consideration has always been the students, and we would not have persisted along a course that would impede their ability to pursue higher education. If the tests were an essential part of the admissions process, then they would adapt, and they did. Our goal has always been to improve the product and the process by which crucial determinations of selection are made. We believed both aspects of our goal, that is improved product and process, would be achieved through simple disclosure. We have never stated nor even implied that Government should play a role in the development or analysis of the content of any of these exams. Rather through disclosure we would permit the experts to debate this issue among themselves and allow the industry to improve itself.

We were told originally that there was a finite number of quality questions for these tests. Then later we were told that although there were many possible questions it took time to construct them and we were warned that haste makes waste. Our position, however, was that scrutiny produces the best product. Let me simply quote E. Belvin Williams, the senior vice president for testing at ETS. ETS, incidentally, as you recall, was one of our leading opponents. Mr. Williams stated in a Times article that diclosure produced a "health reexamination" of the organization's internal procedures for developing and administering exams. On the issue of test quality, he added, "I can say with absolute certainty that the quality did not go down on any of our tests. To the contrary, it is better than ever."

Again, our only concern is for the primary consumer of this product, that is the students. As stated in New York Times editorial in

March of this year, although the emphasis on college entrance examinations is declining in colleges across the Nation, they still play an important role in the futures of more than 1 million students a year. College admissions may become better for this change. Most important, the students who take the tests will feel they're fairer.

I am here today to state that test disclosure is working well in New York and that enactment of a Federal law, that is H.R. 1662, would at least finalize the issue of disclosure which now appears to have few real enemies. It is important that there be a single standard throughout the country to which the test agencies must adhere. It is important to pass this bill to put an end to the lobbying that is otherwise forced to go on in every State that considers similar legislation. Such lobbying, of course, is paid for with student fees and that—and the fact that better use could be made of this money is indeed obvious.

I said that there were few real enemies to the issue of disclosure. Now I say this in light of the industry changes that have occurred since the enactment of our bill in 1979. Within the first few months, the boards which administer the law school admissions test, the graduate management admissions test, and the graduate record exam decided to apply New York's disclosure requirements nationally. They believed that students deserved this service regardless of their address, and they realized that spreading the cost among all the students would reduce the price increase they felt was necessary to impose. Therefore, this change has been in effect for two years now. College Board, who administers the PSAT and SAT, made a similar decision about six months ago. Coincidentally, their action came very shortly after the discovery by a few high school students of errors on the PSAT and SAT, errors I might add that when corrected made new students eligible for scholarships ranging from totals of $1,000 to $8,000. The importance of that discovery to those students is also obvious.

These tests, the SAT, GRE, GMAT, and LSAT comprise the vast majority of this market and they have all moved to disclosure nationally. It has now become an accepted practice and it should be codified with this legislation. If all people and industries were permitted to act voluntarily, as the test agencies appear to be asking you now to allow them to do, then there would be no laws. Preserve this practice of disclosure. I ask that you enact H.R. 1662.

I would now like to focus my specific remarks on 1661 to bill section 7, entitled testing costs and fees to students. Most importantly, this section requires the test agencies to annually disclose information regarding the volume of tests given and the fees collected for those tests as well as the expenses incurred for the development of these tests. An analysis of this information is vital to determining what students are actually paying for when they register to take these exams. That is, we believe that the fees exceed the cost of the service.

Test agencies do not appear to be concerned with the price of their product. After all, most students are compelled to take these exams for admissions to programs in higher education, and in most cases, there is only one test available. Without competition where is the incentive to economize?

We have never received data to refute our contention, for example, that only 5 percent of the students' test fees used to develop the SAT question. Now, this morning we were told that it was 9 percent. Even 9 percent is something that does not bear out their responses as was being brought out by the chairman. We are, therefore, distressed by the excessive fee increases as a result of disclosure legislation. Disclosure of the information required in H.R. 1662 would explain all of this.

I would also like to add at this point that we are delighted to learn that only inflation is causing the increase in these costs and these fees since up to now we've been blamed by passing this legislation for these increases, and this morning we were told that by virtue of the inflation rate they should be higher than they are now. This is the first that I've heard of this, and as I say, we're delighted to learn of it. I would like to add that when they were including—when they were stating the different factors that go into the 9 percent that was used for test development when they were trying to analyze those pieces or those activities of the test agencies that are relevant to disclosure legislation, I don't believe they included their lobbying costs which in New York exceeded $20,000. I think that's another piece of information that ought to be included when they give you the list of activities relevant to the disclosure legislation.

As legislators, your constituents will ask you why these fees are increasing. Before you accept the industry's explanation, you have a right to see the facts.

In closing, let me again state that Federal legislation is necessary for the continuation of test disclosure. Voluntary disclosure for the Nation was done only after disclosure was required in New York State, and furthermore, as you all know, what is given voluntarily can easily be taken away.

Thank you for the opportunity to appear today, and again let me extend Senator Lavalle's apologies for his unavoidable absence.

[Prepared statement of Mary Ann Austen follows:]

PREPARED STATEMENT OF MARY ANN M. AUSTEN COUNSEL TO NEW YORK STATE SENATOR KENNETH P. LAVALLE

MR. CHAIRMAN, MY NAME IS MARY ANN AUSTEN AND I SERVE AS COUNSEL TO SENATOR KENNETH P. LAVALLE, THE SPONSOR OF NEW YORK'S TRUTH-IN-TESTING LAW. LET ME EXTEND HIS APOLOGIES FOR NOT APPEARING BEFORE THIS COMMITTEE, HOWEVER, HE HAD PREVIOUSLY SCHEDULED MEETINGS WHICH HE COULD NOT REARRANGE.

I APPRECIATE THE OPPORTUNITY TO BE HERE IN HIS PLACE AND I HOPE TO PROVIDE YOU WITH INFORMATION WHICH SHOULD ASSIST IN YOUR DECISION TO REPORT H.R.1662 FAVORABLY OUT OF COMMITTEE.

IT HAS BEEN ALMOST TWO AND A HALF YEARS SINCE TRUTH-IN-TESTING LEGISLATION WAS SIGNED INTO LAW IN NEW YORK. THE DIRE PREDICTIONS OF THE OPPOSITION HAVE NOT BECOME A REALITY; THE TESTS ARE STILL OFFERED, THE STUDENTS STILL TAKE THEM AND THE COLLEGES STILL USE THEM. ALL THE PLAYERS, THAT IS THE TEST AGENCIES, THE COLLEGES AND MOST IMPORTANTLY THE STUDENTS HAVE ACCOMODATED THE CHANGES AND HAVE GONE ON ABOUT THEIR BUSINESS. BY ALL REPORTS, THE ATMOSPHERE IS QUITE CALM. IT IS IMPORTANT TO EMPHASIZE THIS FACT TO YOU, ESPECIALLY TO THOSE OF YOU WHO REMEMBER THE EARLY HEARINGS IN THE SUMMER OF 79 WHEN THE TESTIMONY THREATENED CERTAIN CHAOS FROM THE DEMISE OF A TIME HONORED SYSTEM. YOU MIGHT SAY THAT WE WHO HAVE ALWAYS SUPPORTED TEST DISCLOSURE HAD MORE FAITH IN THE RESILIENCE OF THE TESTING SYSTEM THAN THOSE WHO RAN IT. WE BELIEVED THE PROCESS COULD SUSTAIN THE NECESSARY CHANGES AND WE, AS YOU NOW KNOW, WERE RIGHT.

OUR FIRST CONSIDERATION HAS ALWAYS BEEN THE STUDENTS AND WE WOULD NOT HAVE PERSISTED ALONG A COURSE THAT WOULD HAMPER THEIR ABILITY TO PURSUE HIGHER EDUCATION. IF THE TESTS WERE AN INTEGRAL PART OF THE ADMISSIONS PROCESS THEN THEY WOULD ADAPT--AND THEY DID. OUR GOAL HAS ALWAYS BEEN TO IMPROVE THE PRODUCT AND THE PROCESS BY WHICH CRUCIAL DETERMINATIONS OF SELECTION ARE MADE. WE BELIEVED BOTH ASPECTS OF OUR GOAL IMPROVED PRODUCT AND PROCESS WOULD BE ACHIEVED THROUGH SIMPLE DISCLOSURE. WE HAVE NEVER STATED NOR EVEN IMPLIED THAT GOVERNMENT SHOULD PLAY A ROLE IN THE DEVELOPMENT OR ANALYSIS OF THE CONTENT OF ANY OF THESE TESTS. RATHER THROUGH DISCLOSURE WE WOULD PERMIT THE EXPERTS TO DEBATE THIS ISSUE AMONG THEMSELVES AND ALLOW THE INDUSTRY TO IMPROVE ITSELF.

WE WERE TOLD ORIGINALLY THAT THERE WAS A FINITE NUMBER OF QUALITY QUESTIONS FOR THESE TESTS. THEN WE WERE TOLD THAT ALTHOUGH THERE WERE MANY POSSIBLE QUESTIONS IT TOOK TIME TO CONSTRUCT THEM AND WE WERE WARNED THAT "HASTE MAKES WASTE." OUR POSITION HOWEVER WAS THAT SCRUTINY PRODUCES THE BEST PRODUCT. LET ME SIMPLY QUOTE E. BELVIN WILLIAMS THE SENIOR VICE PRESIDENT FOR TESTING AT ETS, (INCIDENTLY AT THE OUTSET ETS WERE THE LEADING OPPONENTS ON THIS ISSUE). MR. WILLIAMS STATED IN A N.Y. TIMES ARTICLE THAT DISCLOSURE PRODUCED "A HEALTHY RE-EXAMINATION" OF THE ORGANIZATIONS INTERNAL PROCEDURES FOR DEVELOPING AND ADMINISTERING EXAMS. ON THE ISSUE OF TEST QUALITY HE ADDED, "I CAN SAY WITH ABSOLUTE CRETAINTY THAT THE QUALITY DID NOT GO DOWN ON ANY OF OUR TESTS. TO THE CONTRARY, IT IS BETTER THAN EVER, (N.Y. TIMES NOVEMBER 23, 1980).

AGAIN OUR ONLY CONCERN IS FOR THE PRIMARY CONSUMER OF THIS PRODUCT, THAT IS THE STUDENTS. AS STATED IN A NEW YORK TIMES EDITORIAL IN MARCH, 1981:

> Although the emphasis on college entrance examinations is declining in colleges across the nation, they still play an important role in the futures of more than a million students a year. College admissions may become better for this change. Most important, the students who take the tests will feel they're fairer (N.Y.Times, March 20, 1981).

I AM HERE TODAY TO STATE THAT TEST DISCLOSURE IS WORKING WELL IN NEW YORK AND THAT INACTMENT OF A FEDERAL LAW, HR. 1662 WOULD AT LAST FINALIZE THE ISSUE OF DISCLOSURE WHICH NOW APPEARS TO HAVE FEW REAL ENEMIES. IT IS IMPORTANT THAT THERE BE A SINGLE STANDARD THROUGHOUT THE COUNTRY TO WHICH THE TEST AGENCIES MUST ADHERE. IT IS IMPORTANT TO PASS THIS BILL TO PUT AN END TO THE LOBBYING THAT IS OTHERWISE FORCED TO GO ON IN EVERY STATE THAT CONSIDERS SIMILAR LEGISLATION. SUCH LOBBYING IS OF COURSE PAID FOR WITH STUDENT FEES AND THAT BETTER USE COULD BE MADE OF THIS MONEY IS INDEED OBVIOUS.

I SAID THAT THERE ARE FEW REAL ENEMIES TO THE ISSUE OF DISCLOSURE. I SAY THIS IN LIGHT OF THE INDUSTRY CHANGES THAT HAVE OCCURED SINCE THE ENACTMENT OF OUR BILL IN 1979. WITHIN THE FIRST FEW MONTHS, THE BOARDS WHICH ADMINISTER THE LSAT, GMAT AND GRE DECIDED TO APPLY NEW YORK'S DISCLOSURE REQUIREMENTS NATIONALLY. THEY BELIEVED STUDENTS DESERVED THIS SERVICE REGARDLESS OF THEIR ADDRESS AND THEY REALIZED THAT

SPREADING THE COST AMONG ALL THE STUDENTS WOULD REDUCE THE PRICE INCREASE THEY FELT WAS NECESSARY TO IMPOSE. THEREFORE, THIS CHANGE HAS BEEN IN EFFECT FOR TWO YEARS NOW. COLLEGE BOARD, WHO ADMINISTERS THE PSAT AND SAT, MADE A SIMILAR DECISION. THEIR ACTION CAME VERY SHORTLY AFTER THE DISCOVERY BY A FEW HIGH SCHOOL STUDENTS OF ERRORS ON THE PSAT AND SAT. ERRORS I MIGHT ADD, THAT WHEN CORRECTED MADE NEW STUDENTS ELIGIBLE FOR SCHOLARSHIPS RANGING FROM TOTALS OF $1000 to $8000. THE IMPORTANCE OF THAT DISCOVERY TO THOSE STUDENTS IS OBVIOUS.

THESE TESTS, THE SAT, GRE, GMAT AND LSAT COMPRISE THE VAST MAJORITY THIS MARKET AND THEY HAVE ALL MOVED TO DISCLOSURE NATIONALLY. IT HAS NOW BECOME AN ACCEPTED PRACTICE AND IT SHOULD BE CODIFIED WITH THIS LEGISLATION. IF ALL PEOPLE AND INDUSTRIES WERE PERMITTED TO ACT VOLUNTARILY -- AS THEY APPEAR TO BE ASKING YOU TO DO -- THEN THERE WOULD BE NO LAWS. PRESERVE THIS PRACTICE OF DISCLOSURE; I ASK THAT YOU ENACT H.R. 1662.

I WOULD NOW LIKE TO FOCUS MY SPECIFIC COMMENTS ON H.R. 1662 TO BILL SECTION SEVEN ENTITLED, "TESTING COSTS AND FEES TO STUDENTS."

Mr. **Weiss**. Ms. Austen, thank you very much for your testimony and for coming down to testify on behalf of Senator Lavalle.

Had New York considered including the equivalent of section 7 in its legislation?

Ms. **Austen**. At the time we were doing this, that seemed to be, although it was one aspect of the total amount of information we were trying to obtain, we felt that it was such a new issue that we did not want to cloud the issue with so many aspects, and in fact our goal at the time, which as it was, was improved testing information to students about the questions themselves and the reports and the statistics done on them, we didn't want to, as I say, cloud that with this other information which they would have used, they meaning the test agencies, in their lobbying efforts. We felt, perhaps the detriment of our main goal so we excluded it. It's something that Senator Lavalle is extremely concerned with now because of the increasing test fees. With everything else increasing, this is one additional thing that his constituents and everyone else would like the burden reduced on, and it's something that he's seriously looking at now in New York.

Mr. **Weiss**. Well, it must be gratifying to you to find that a piece of legislation which originally was objected to so violently now is being cited as being so effective as to preclude any further legislation in the field here or any place else.

Ms. **Austen**. Wonderful. It's quite a different tune.

Mr. **Weiss**. I gather that there are still some tests which are not being offered or to which the legislation does not apply because of legal action.

Ms. **Austen**. Yes.

Mr. **Weiss**. Could you touch on that briefly?

Ms. **Austen**. Yes; there are, I believe, six exams. There was a list compiled by the State Education Department upon the passage of this legislation of tests offered in New York that fell under the definition in our bill of tests for admission to postsecondary institutions, and the list was approximately 26 tests in length. Originally, the State Education Department announced that 20 of them were leaving, leaving only 6, the 6 which most of us think of, as a matter of fact they are the ETS tests but in any case with the passage of some legislation—we were unaware of some of these tests existence. They're very, very small tests. Sometimes one test is made up by an individual, for instance one is offered to some nursing students that other larger tests—tests that are available take care of in other parts of the State. They were a regional and much more informal activity.

We passed legislation when we found out about some of these in their size. We included which was a portion of your first bill that was an exemption for small tests. We included an exemption for small tests, however, requiring them every 3 years to put one of their tests on file so that there was something for some sort of disclosure of their tests, however recognizing the burden of a 500, for example, person program as opposed to the 200,000 SAT program most of them have come back. There are a few that have not come back for one reason or another. It's not within my knowledge at this point to know if, in fact, they're given anywhere else anymore. They may just not exist anymore. In any case, the slack has been picked up by other tests. Apparently we have not had any complaints about students wanting a program but not being able to take an exam and having to go to another State. The one exam that is not—that is given, however, not subjected to the law in any respect is the MCAT which is the test for admission to medical school, and that is as a result of an injunction that was obtained by the American Association of Medical Colleges in January of 1980, and that is still—has still not been determined on the merits. It was an injunction. The issue is copyright, and, as I stated in my testimony last year, the test for injunctions in copyright matters is much less demanding than it is in many other areas because of the nature of copyrights so the issue of whether or not there is a copyright violation here has not yet been determined, and we submit continuously and especially in light of the time that has gone by and the fact that everyone has adjusted so well evidenced by the testimony today. This is right up there with motherhood and apple pie at this point seemingly. They still contend that, assuming that they haven't dropped the suit, that they cannot comply, that they are still burdened by the finite number of quality questions that they always stated which was an argument that ETS posed originally in spite of what they're saying now, and also it's interesting to note that the dental admissions tests which is a similar test in the sense that it is composed primarily of mathematics and science has been disclosed in New York. They don't have a problem, in fact, they are exempt by virtue of how many students take the test, but in spite of that, they're disclosing anyway. They seem to be able to construct a test of that nature which seems to suit their purposes in terms of all testimony that came out when they originally testified about wanting to curtail the kind of student that couldn't make it all the way through. Their test accomplishes that goal which is apparently also a goal of the medical schools. They can disclose, but the medical schools contend they still cannot. I just wanted to point that out. It's still pending.

Mr. **Weiss**. Thank you very much and again our appreciation too. Mr. Williams.

Mr. **Williams**. Thank you, Mr. Chairman. Ms. Austen, we understand that the New York Act has been amended since it was first adopted. Can you tell us generally the nature of those amendments?

Ms. **Austen**. Well, one of them was the exemption for small volume tests which, as I say, was a portion of the section rather of the original bill.

Mr. **Williams**. Is that a major amendment?

Ms. **Austen**. Yes, it was. It permitted testing to continue that we felt, as I say, was overburdened by this regulation. The precedent exists for exclusion exemptions for small businesses in many other kinds of regulations. It's something that occurs all the time. We're often asked if it's good for one, why isn't it good for all. Well, the fact is, its smallness sometimes counts for something. It was that exemption.

There was also a clarification that the bill only—or the law would only apply to tests given in New York rather than used in New York. There was some question the College Board brought suit—that if a test was given to a student in Idaho, for example, who applied to Columbia in New York, that if Columbia used that exam because college board was giving a different test in different States—they were giving one SAT test in New York and a different one in the rest of the country—the test that was used by the admissions officer in New York would come under the purview of the legislation and therefore would have to be disclosed and would spoil their—their action to give a different test in other States.

There was flexibility given in the law for the special tests that are given for handicapped students, walk-in students, makeups, and Sabbath observers that created some flexibility to allow the test agencies to run a particular exam a few times because of the low volume of the takers to use that for a couple of years and then disclose it. I believe that was the——

Mr. **Williams**. As you know, there are those of us who are concerned that testing, and by the way professional licensing procedures also, are hurdles that may be overly high and prevent significant segments of the population from achieving their educational and professional goals. Many of us, and the testing community, I think, is among them are concerned that economic advantage leads to socioeconomic advantage, leads to scholarship advantage, leads to college, which leads to economic advantage and the tests may not reflect that cycle. They may speed it up. How does the Truth in Testing Act in New York assist in slowing down that cycle and lowering those hurdles?

Ms. **Austen**. Well, I think it did nothing to accomplish that goal immediately. It began a process which over the years we hope will begin to address itself to that problem. There were a couple of things. First, we felt that by exposing the exams to all students rather than the students who take these coaching courses that in effect had the exams for $300—in Long Island now we have students taking the coaching courses for the PSAT at $250. It has gotten totally out of hand but we felt that by putting them on file and now what ETS is doing you don't have to go—our bill permitted you to go to the Commissioner of Education to get the exam if you hadn't taken the exam yourself—the retired exam. ETS has, to my knowledge, has opened their files. If they have disclosed it anywhere, it's available to anyone or simply writing in. So by giving that, by making available other tests without having to pay admission to a course, we felt that students who could take advantage of just preparing themselves by not learning new things, just looking at the exams themselves and perhaps having someone do on an informal basis the coaching—a teacher perhaps or a parent or whomever—we felt that that would begin to close that gap, that we would not have the economic barrier to a coaching setting that existed.

There was also, it came to our attention, a geographic barrier. If you didn't happen to live near a center in a metropolitan area that offered these courses, you couldn't take them no matter how much you had so there were those two barriers. We felt we've begun to close that gap.

Second, we felt that all this information that was coming out about the test that was again pointed out earlier this morning, the coachability of the tests and some of these other things that they are so willing to give out now were not available a couple of years ago. People that assisted in these studies were unable to get some of the information so we felt that reports that would come out as a result of this new openness would show admissions people, we can't change how admissions occur legislatively, we felt. There has to be so much flexibility that that wasn't something that we felt could be regulated, but we did feel that admissions officers were blindly using these tests because they felt that they did so much. They could determine who was a qualified student and who wasn't.

At our hearing we had admissions officers parrot exactly testimony that had been given to them by testing agencies. The relationship had occurred over the years and they were blind believers, and I want to tell you that they are somewhat embarrassed at this point and there are quotes that I have that have been reported in major newspapers and so forth. When these errors came out in the exams, these people were astonished. They were supposed to be infallible, not only as to the correct answers but also as to what they predicted, and they're not and the admissions people are now—they have to in good conscience think back what am I really using this tool for and how am I going to use it in the future, and we feel that that again change over the years will change the process but it's not an immediate change.

Mr. **Weiss**. Ms. Austen, you might be—I'm sure that you're aware of a group whose executive director testified before us in July, the East Harlem College and Career Guidance Center. They operate in East Harlem and because of the availability of the tests now, they're able at no cost at all to set up a program for kids in the East Harlem community who before wouldn't have thought about taking the SAT because they were just convinced that they could never pass those examinations so that, in fact, just sheer disclosure is having, I think, some very profound societal effects, and I think that your analysis is quite accurate.

Again, thank you very, very much and we hope that the next time we invite you down we'll be a little closer to having Federal legislation on the books.

Ms. **Austen**. That could be the bill signing ceremony.

Mr. **Weiss**. Right. Well, I suspect you're going to have to testify before the Senate before that happens.

Ms. **Austen**. OK.

Mr. **Weiss**. Our next and last witness for the day is Mr. John Cooper, president, Association of Medical Colleges. Mr. Cooper.

STATEMENT OF JOHN COOPER, PRESIDENT, ASSOCIATION OF MEDICAL COLLEGES

Mr. **Cooper**. Thank you very much Mr. Weiss. The association appreciates this opportunity to report on its continuing experience with legislation and legislative proposals relating to standardize testing.

I am Dr. John A. D. Cooper, president of the Association of American Medical Colleges. It was felt so important for me to be here today that I left our association's annual meeting of some 3,900 medical educators to do so. We hope our testimony helps the subcommittee to determine that the purposes of these hearings have been accomplished and that further consideration of this bill is not warranted. We strongly believe this to be a rational conclusion from the information presented at these hearings and from the accumulated nationwide experience during the interval since this legislation was first introduced.

We would like to review briefly the reasons why we are told that this legislation is necessary.

First, there is widespread concern in the States about this matter.

Second, we can expect a "proliferation of State-level disclosure laws that will create a complex regulatory web of for test givers."

Third, test sponsors are not sufficiently accountable.

Fourth, test disclosure is necessary to detect and prevent scoring errors.

Fifth, test disclosure is necessary to equalize access to coaching courses.

We believe that none of these reasons are valid reasons for legislation either at the Federal or State level.

Part of the response to the question of widespread concern was offered by Barbara Lerner in her article in the March 1981 issue of American Psychologist. She reviewed the data from 10 different public opinion surveys. In none of the 10 was she able to find evidence of any widespread concern that proponents would have us believe exists.

But what has been happening at the State level? Mr. Perkins began these hearings over 2 years by cautioning those present that the issues under consideration were traditionally the province of the States. Directly related to the issues raised by Mr. Perkins, testing legislation has been studied quite explicitly in almost one-third of the State legislatures during the past 2 years. The result to date has been most instructive. Not a single State has passed legislation in this area since hearings on this subject were begun by the U.S. Congress. The message from this accumulated experience should be unmistakable. It should at least signal very clearly that there is a great amount of doubt in a great many States about the wisdom of such legislation.

There is overwhelming evidence from State-level activity also easily disposes of another phantom—the second reason for legislation, that is the "proliferation of State-level disclosure laws that will create a complex regulatory web for test givers." Recent amendments to existing New York and California law do not alter this conclusion.

A very important finding of these State inquiries has been the repeated recognition of the diversity and complexity of the issues being addressed. It was recognized that one test is not pretty much like another and that the purposes of tests vary widely. We have attempted over a number of years to bring this point to the attention of those who would have such legislation including Dr. Sunder Lavale and Ms. Austin.

Our test does differ from the other test and as a matter of fact that was recognized by Representative Marks in California in his amendment and we were exempt from that amendment on the basis that it is different. We do use a different procedure than are used by the other test designers for the construction and the validation of the test.

We spent a lot of money to develop what we thought was a much more effective test. tests which gave much better information to those who took it and those who were using the test results in anyway. We think that has been accomplished and we want to continue to make the point that we are different.

The circumstances under which the tests are administered and the ways in which test scores are used were found to be variable. In short what may have appeared appropriate and equitable to require one program was seen to be inappropriate and destructive for another. Our medical college admission test is the case in point. We have argued that the MCAT is a highly specialized achievement examination with specifications drawn very tightly to insure maximum relevance to medical education and to provide maximum guidance to the candidate for his or her preparation for the test. We did disclose the nature of those restrictions for any legislation that was introduced in either California, New York, or proposed in the national Congress.

The MCAT manual which the students can obtain at a very low cost provides them with very explicit instructions on the areas in which the test items are drawn. And second, it provides them with a test which permits them to and how the individual types of questions are developed. It also provides them with a test which they can use themselves in trying to determine their preparation, their readiness to take a test of this kind. Data were collected from a cross-section in developing the questions. Data were collected from a cross-section of approximately 150 medical faculty to determine what they considered as relevant to the study and practice of medicine. Testimony was heard last summer from an MCAT examinee, Ms. Carolyn Bennett, that she and her colleagues considered the MCAT a fair representation of what students should know to get into medical school.

All of this is now in serious jeopardy as a result of this bill. Disclosure of test questions and answers as required by the New York law and by H.R. 1662 would destroy the MCAT since under these conditions there simply is not a sufficient supply of questions that would meet the specifications of the test. Disclosure is also incompatible with the conditions under which the MCAT must be administered with the result that our ability to write and test new questions and equate new forms would be seriously compromised.

These kinds of considerations were apparently persuasive in several States since we were advised a number of times that should legislation prove feasible, the MCAT would be excluded. The same issues, by the way, were fundamental in our pleadings before the Federal court in the northern district of New York that led to our being awarded a preliminary injunction preventing New York from enforcing the disclosure provisions of its law against the MCAT.

The third issue raised dealt with the accountability of test sponsors. I have explained in previous testimony how the MCAT program is directly accountable to the 126 medical schools of the United States, their faculty, students, and administration and how we provide for the continuing counsel of the candidate's premedical faculty. Proponents direct the issue of accountability not only to test contents but also propose extending disclosure to all studies, data analyses, reports, and so forth as though test sponsors control all research and evaluation associated with the test and have no reason to feel accountable to the public which it serves. On the occasion of our last appearance in this forum, we submitted three early studies conducted independently of the AAMC as a demonstration that the test sponsor is not the sole source of information about a test. We have now counted 21 independently prepared studies relating to the interpretation and use of the new MCAT that have appeared in the literature. Appendix A includes 21 of these citations for your perusal. The list cites studies conducted at 15 different schools. In addition, the AAMC has developed a cooperative arrangement with other schools to stimulate their interest in conducting local studies that will contribute at the same time to nationally aggregated information. The result is that over 40 schools are actively engaged in studying the effectiveness of the MCAT and will be preparing articles for publication that the test sponsor has no opportunity to control. Contrary to the claims of the bill's proponents, it seems that a high degree of accountability exists in this situation where approximately one-third of the institutional users of a test are actively involved in the evaluation of its effectiveness.

The fourth reason cited suggested that the test disclosure was necessary to detect and prevent scoring errors. The term "scoring errors" as we have understood it have been used in at least two ways: One, to refer to an incorrectly calculated score, and two, to refer to a question that has ambiguities in the correct solution that has been identified. It is our contention that the exposure of test questions and answers as required by H.R. 1662 is a classic case of killing fleas with a sledge hammer. If a serious problem exists in this regard, other review mechanisms preserving the security of the examination materials are certainly available to provide the necessary assurances. In the MCAT program we have instituted a series of checks and balances both before and after the administration of every test in order to achieve the highest probability of detection of an error of either kind. We have detailed in previous testimony all of the steps taken in the development and testing of each individual question. In addition to these safeguards common to the industry as a whole, we have special field tryouts before a question is ever used where samples comparable to our examinee population are engaged in a dialog seeking a variety of reactions to questions under development. And that sample does include minorities among others. In addition, after each administration we have an elaborate formal score verification process that is designed to be a final check on possible problems in the scoring of an answer sheet, questions with potential ambiguities in the keying of the correct answer, and the potential errors in the determination of scaled scores and their equivalency with previous forms. These checks employ visual inspection of answer sheets, comments from examinees who have taken the test and those comments are immediately forwarded to the test developers and are acted upon, a review of item characteristics, and sophisticated statistical techniques. Though we are confident that we have instituted every reasonable precaution against errors, we continue to evaluate our system through the ongoing use of teams of external consultants who are recognized as leaders in their field.

And with your permission I would just like to give their qualifications so that you can evaluate the kind of people that we're using. Dr. Robert Lynn, professor of education, University of Illinois, who is president of the National Council on Measurement in Education. Dr. Laurie Shepherd, professor of education, University of Colorado, president-elect of the National Council on Measurement in Education, former editor of the Journal of Educational Measurement. Richard Jaggar, professor of education, the University of North Carolina in Greensboro, former editor of the Journal of Educational Measurement.

The question then is what would the exposure of all test questions and answers contribute? The answer is that it would not only destroy the current test specifications at least for the MCAT, but would result in the need to produce greater quantities of material with a corresponding loss in available resources to maintain the current checks on question quality. In short, the proponents of this legislation are in one sense in an envious position. For test sponsors, on the other hand, it is a Catch-22. If disclosure is mandated, more poorer questions will result and more scoring errors will be identified, justifying, of course, the imposition of the law to begin with and the intrusion of the Federal Government into the educational matters that are by long tradition the responsibility of the private sector.

A fifth reason we have heard cited in support of testing legislation is the need to equalize access to coaching courses. This seems to be a major justification since considerable attention is planned for the subject tomorrow at hearings here. Several subquestions have been raised or implied in the discussion of coaching courses. First, can they affect test performance, how much, and is that bad? Second, if some can increase performance, should the U.S. Congress attempt to equalize access to them? Third, how does H.R. 1662 address the issue?

Concerning the impact of test preparation courses that they can have on test performance, we have been trying to collect some empirical data about participation in such experiences and their association with changes and performance on the MCAT. The deeper we probe the area, the more complex, the more contaminated with uncontrollable factors the question becomes, and the more hazardous any useful generalizations seem. For example, our data is volunteered by the candidates. We do not have access to the list of people that are taking the coaching courses. It reflects a wide disparity of review experiences, the participants have differing levels of motivation, they start with different backgrounds, et cetera. With these caveats, we are tentatively concluding that though on the average examinees acknowledging participation in commercial review courses exhibit gains in performance on a second administration of the MCAT, the gain is only slightly greater than that observed for those repeating the test after the same interval, but not reporting participation in a formal course. The magnitude of the average incremental gain is on the order of one-tenth to one-half of a scaled score point. The magnitude of the change associated with the group is not reporting participation in a commercial review course is on the order of one-fourth to one scaled point. To report more specific figures at this time we think would imply a stability and the confidence for these results that we have not yet been able to establish. We did find consistently that the higher gains were for the content specific sections of the test, that is biology, chemistry, and physics and that the gains observed at the lower ends of the ranges were associated with skills analysis, reading, and quantitative tests. As related initially at the lower levels on the Sills Analysis: Reading and quantitative subtests gained least from the typical review course experience. Since these test preparation experiences tend to be heavily content oriented rather than emphasizing the development of basic thinking skills, such a test preparation experience is likely to have relatively little payoff for examinees with less well developed thinking skills.

If as it now appears there may be an effect on MCAT scores associated with participation in a review course, is that bad? Changes in performance on the MCAT as a result of a test preparation course are not a threat to the validity of our test. In fact, to the extent that a test preparation course or any other effort that the student uses to increase and gain knowledge enhances the candidate's knowledge or skills measured by the test, the MCAT would be invalid if it did not reflect that improvement. In this regard, it should be understood that we regularly monitor the leakage of test materials from the program to assure ourselves that materials reused on subsequent forms have not been compromised.

In summary, none of these data have caused us to harbor concerns about the validity of the test. The magnitude of the changes within reasonable bounds, and the type of change and interaction among types of performance even offer evidence for the construct validity of the test, that is the diagnostic value of the skills subtests. We are continuing our research and will report specific findings as we establish the stability of our earlier indications.

The second issue related to coaching was whether the U.S. Congress should try to equalize access to test preparation courses or for that matter to any resource that might be effective in improving performance. Proponents simply—proponents imply that it should—but the issue has not been directly confronted. Should the answer be yes, then new legislation will be required.

This brings us to the last issue which is whether H.R. 1662 addresses the concern about equalizing access. Presumably the link is in the increased availability of test questions and answers resulting from mandatory disclosure. In the comments accompanying the introduction of H.R. 1662, this statement is made and I quote:

> And it (the testing legislation) would lessen inequities among students created by expensive coaching schools by giving everyone equal access to information about the test and the questions themselves—not widen the gap between students.

At best, this is an unproven assertion. At worst, testing legislation will exacerbate the problem in the sense of the above quote in that it will widen the gap.

First, from the experience of the college boards in New York, we know that the small numbers requesting their questions and answers—fewer than 5 percent—disproportionately more are from the higher socio-economic groups and from those already achieving high scores on the test. If there is a benefit to having access to questions and answers, whom then will it benefit most? Clearly, not the disadvantaged. If the test preparation experiences continue to be expensive and therefore access for the disadvantaged continues to be limited, we see nothing in H.R. 1662 that will change that. And if there is any value to such experiences, advantaged groups in the population will continue to be the primary beneficiaries. It also seems quite reasonable that putting more information into the hands of the operators of such courses will only enhance their potential to offer effective programs of preparation. It is predictable also that test preparation courses may be made more attractive by their being perceived as the collector and organizer of the flood of questions and answers that are suddenly available. Unexplained, uninterpreted, unorganized bursts of questions and answers will only succeed in increasing dependence on external support while satisfying the resource requirements of the operators in the test preparation business to make money.

We fail to see how these arguments advance the support of H.R. 1662 survive the careful scrutiny of evidence or logic. This intervention would be contrary to the clearly expressed views of the people that Federal presence should be reduced. With the severe restriction in available resources it does not seem logical to spend the public purse on unnecessary regulatory initiatives. In our view these hearings have amply demonstrated all the fundamental flaws of H.R. 1662 and we urge that the subcommittee not report this measure to the full committee. Thank you very much for your kind attention.

[Prepared statement of John Cooper follows:]

PREPARED STATEMENT OF JOHN COOPER, PRESIDENT, ASSOCIATION OF AMERICAN MEDICAL COLLEGES

Chairmen Perkins and Simon, and Members of the Subcommittees on Elementary, Secondary, and Vocational Education, and on Postsecondary Education:

The Association of American Medical Colleges (AAMC) appreciates this opportunity to report on its continuing experience with legislation and legislative proposals relating to standardized testing. I am Dr. John A. D. Cooper, the President of the AAMC. I felt it was so important to be here today, that I left our Association's Annual Meeting to do so. We hope our testimony helps the subcommittees determine that the purposes of these hearings have been accomplished and that further consideration of this bill is not warranted. We strongly believe this to be a rational conclusion from the information presented at these hearings and from the accumulated nationwide experience during the interval since this legislation was first introduced.

We would like to review briefly the reasons why we are told that this legislation is necessary.

1. There is widespread concern in the states about this matter.
2. We can expect a "proliferation of state-level disclosure laws that will create a complex regulatory web for test givers."
3. Test sponsors are not sufficiently accountable.
4. Test disclosure is necessary to detect and prevent scoring errors.
5. Test disclosure is necessary to equalize access to coaching courses.

Presented by Dr. John A. D. Cooper, President, Association of American Medical Colleges, before the House Subcommittees on Elementary, Secondary, and Vocational Education and Postsecondary Education, Committee on Education and Labor, November 4, 1981.

Part of the response to the question of widespread concern was offered by Barbara Lerner in her article in the March 1981 issue of the American Psychologist. She reviewed the data from 10 different public opinion surveys. In none of the 10 was she able to find evidence of the widespread concern proponents would have us believe.

But what has been happening at the state level? Mr. Perkins began these hearings over two years ago by cautioning those present that the issues under consideration were traditionally the province of the states. Directly related to the issue raised by Mr. Perkins, testing legislation has been studied quite explicitly in almost one-third of the state legislatures during the past two years. The result to date has been most instructive. Not a single state has passed legislation in this area since hearings on this subject were begun in the U.S. Congress. The message from this accumulated experience should be unmistakable. It should at least signal very clearly that there is a great amount of doubt in a great many states about the wisdom of such legislation.

This overwhelming evidence from state level activity also easily disposes of another phantom — the second reason for legislation, i.e. the "proliferation of state-level disclosure laws that will create a complex regulatory web for test givers." Recent amendments to existing New York and California law do not alter this conclusion.

A very important finding of these state inquiries has been the repeated recognition of the diversity and complexity of the issues being addressed. It was recognized that one test is not pretty much like another and that the purposes of tests vary widely. The circumstances under which they are administered and the ways in which test scores are used were found to be quite variable. In short what may have appeared appropriate and equitable to require of one program was seen to be inappropriate and destructive for another. Our Medical College Admission Test (MCAT) is a case in point. We have argued that the MCAT is a highly specialized achievement examination with specifications drawn very tightly to insure maximum relevance to medical education and to provide maximum guidance to the candidate for his or her preparation for the test. Data were collected from a cross-section of approximately 150 medical faculty to determine what is relevant to the study and practice of medicine. Testimony was heard last summer from an MCAT examinee, Ms. Carolyn Bennett, that she and her colleagues considered the MCAT a fair representation of what students should know to get into medical school.

All of this is now in serious jeopardy as a result of this bill. Disclosure of test questions and answers as required by the New York law and by H.R. 1662 would destroy the MCAT since under those conditions there simply is not a sufficient supply of questions that would meet the specifications of the test. Disclosure is also incompatible with the conditions under which the MCAT must be administered with the result that our ability to write and test new questions and equate new forms would be seriously compromised.

These kinds of considerations were apparently persuasive in several states since we were advised a number of times that should legislation prove feasible, the MCAT would be excluded. The same issues, by the way, were fundamental in our pleadings before the federal court in the Northern District of New York that led to our being awarded a preliminary injunction preventing New York from enforcing the disclosure provisions of it law against the MCAT.

The third issue raised dealt with the accountability of test sponsors. I have expiained in previous testimony how the MCAT program is directly accountable to the 126 medical schools of the United States, their faculty, students, and administration and how we provide for the continuing counsel of the candidates' pre-medical faculty. Proponents direct the issue of accountability not only to test contents but also propose extending disclosure to all studies, data analyses, reports, etc. as though test sponsors control all research and evaluation associated with the test and have no reason to feel accountable to the public it serves. On the occasion of our last appearance in this forum, we submitted three early studies conducted independently of the AAMC as a demonstration that the test sponsor is not the sole source of information about a test. We have now counted 21 independently prepared studies relating to the interpretation and use of the New MCAT that have appeared in the literature. Appendix A includes 21 of these citations for you perusal. The list cites studies conducted at 15 different schools. In addition, the AAMC has also developed a cooperative arrangement with other schools to stimulate their interest in conducting local studies that will contribute at the same time to nationally aggregated information. The result is that over 40 schools are actively engaged in studying the effectiveness of the MCAT and will be preparing articles for publication that the test sponsor has no opportunity to control. Contrary to the claims of the bill's proponents, it seems that a high degree of accountability exists in this situation where approximately one-third of the institutional users of a test are actively involved in the evaluation of its effectiveness.

The fourth reason we cited suggested that test disclosure was necessary to detect and prevent scoring errors. The term "scoring errors" as we have understood it has been used in at least two ways: 1) to refer to an incorrectly calculated score and 2) to refer to a question that has ambiguities in the correct solution that has been identified. It is our contention that the exposure of test questions and answers as required by H.R. 1662 is a classic case of killing fleas with a sledge hammer. If a serious problem exists in this regard other review mechanisms preserving the security of the examination materials are certainly available to provide the necessary assurances. In the MCAT program we have instituted a series of checks and balances both before and after the administration of every test in order to achieve the highest probability of detection of an error of either kind. We have detailed in previous testimony all the steps taken in the development and testing of each individual question. In addition to these safeguards common to the industry as a whole, we have special field tryouts before a question is ever used where samples comparable to our examinee population are engaged in a dialogue seeking a variety of reactions to questions under development. In addition, after each administration we have an elaborate formal score verification process that is designed to be a final check on possible problems in the scoring of an answer sheet, questions with potential ambiguities in the keying of the correct answer, and the potential errors in the determination of scaled scores and their equivalency with previous forms. These checks employ visual inspection of answer sheets, comments from examinees who have taken the test, a review of item characteristics, and sophisticated statistical techniques. Though we are confident that we have instituted every reasonable precaution against errors, we continue to evaluate our system through the ongoing use of a team of external consultants who are recognized as leaders in their field. The question then is what would the exposure of all test questions and answers contribute? The answer is that it would not only destroy the current test specifications but would result in the need to produce greater quantities of material with a corresponding loss in available resources to maintain the current checks on question quality. In short, the proponents of this legislation are in one sense in an envious position. For test sponsors on the other hand, it is a Catch-22. If disclosure is mandated, more poorer questions will result and more scoring errors will be identified, justifying, of course, the imposition of the law to begin with and the intrusion of the federal government into the educational matters that are by long tradition the responsibility of the private sector.

The fifth reason we have heard cited in support of testing legislation is the need to equalize access to "coaching" courses. This seems to be a major justification since considerable attention is planned for the subject tomorrow. Several subquestions have been raised or implied in discussion of "coaching" courses. First, can they affect test performance, how much, and is that bad? Second, if some can increase performance, should the U.S. Congress attempt to equalize access to them? Third, how does H.R. 1662 address the issue?

Concerning the impact test preparation courses can have on test performance, we have been trying to collect some empirical data about participation in such experiences and their association with changes in performance on the MCAT. The deeper we probe the area, the more complex, the more contaminated with uncontrollable factors the question becomes, and the more hazardous any useful generalizations seem. For example our data is volunteered by the candidates, it relfects a wide disparity of review experiences, the participants have differing levels of motivation, they start with different backgrounds, etc. With these caveats, we are tentatively concluding that though on the average examinees acknowledging participation in commercial review courses exhibit gains in performance on a second administration of the MCAT, the gain is only slightly greater than that observed for those repeating the test after the same interval but not reporting participation in a formal course. The magnitude of the average incremental gain is on the order of one-tenth to one-half of a scaled score point. The magnitude of change associated with the group not reporting participation in a commercial review course is on the order of one-fourth to one scaled point. To report more specific figures at this time would imply a stability for these results that we have not yet been able to establish. We did find consistently that the higher gains were for content specific sections of the test, i.e. Biology, Chemistry, and Physics and that the gains observed at the lower ends of the ranges were associated with the Skills Analysis: Reading and Quantitative tests. A related but very interesting observation suggested that those scoring initially at the lower levels on the Skills Analysis: Reading and Quantitative subtests gained least from the typical review course experience. Since these test preparation experiences tend to be heavily content oriented rather than emphasizing the development of basic thinking skills, such a test preparation experience is likely to have relatively little payoff for examinees with less well developed thinking skills.

If as it appears there may be an effect on MCAT scores associated with participation in a review course, is that bad? Changes in performance on the MCAT as a result of a test preparation course are not a threat to the validity of our test. In fact, to the extent that a test preparation course or any other effort enhances the candidate's knowledge or skills measured by the test, the MCAT would be invalid if it did not reflect that improvement. In this regard it should be understood that we regularly monitor the leakage of test materials from the program to assure ourselves that materials reused on subsequent forms have not been compromised.

In summary, none of these data have caused us to harbor concerns about the validity of the test. The magnitude of the change is within reasonable bounds, and the type of change and interaction among types of performance even offer evidence for the construct validity of the test, i.e. the diagnostic value of the skills subtests. We are continuing our research and will report specific findings as we establish the stability of our early indications.

The second issue related to coaching was whether the U.S. Congress should try to equalize access to test preparation courses or for that matter to any resource that might be effective in improving performance. Proponents imply that it should, but the issue has not been directly confronted. Should the answer be yes, then new legislation will be required.

This brings us to the last issue which is whether H.R. 1662 addresses the concern about equalizing access. Presumably the link is in the increased availability of test questions and answers resulting from mandatory disclosure. In the comments accompanying the introduction of H.R. 1662, this statement is made, "and it [testing legislation] would lessen inequities among students created by expensive coaching schools by giving everyone equal access to information about the test and the questions themselves -- not widen the gap between students." At best, this is an unproven assertion. At worst, testing legislation will exacerbate the problem in the sense of the above quote in that it will widen the gap. First, from the experience of the College Boards in New York, we know that of the small numbers requesting their questions and answers (fewer than 5%), disproportionately more are from the higher socio-economic groups and from those already achieving high scores on the test. If there is a benefit to having access to questions and answers, whom then will it benefit most? Clearly, not the disadvantaged. If test preparation experiences continue to be expensive and therefore access for the disadvantaged continues to be limited -- we see nothing in H.R. 1662 that will change that — and if there is any value to such experiences, advantaged groups in the population will continue to be the primary beneficiaries. It also seems quite reasonable that putting more information into the hands of the operators of such courses will only enhance their potential to offer effective programs of preparation. It is predictable also that test preparation courses may be made more attractive by their being perceived as the collector and organizer of the flood of questions and answers that are suddenly available. Unexplained, uninterpreted, unorganized bursts of question and answers will only succeed in increasing dependence on external support while satisfying the resource requirements of the operators in the test preparation business to make money.

We fail to see how any of these arguments advanced to support H.R. 1662 survive the careful scrutiny of evidence or logic. This intervention would be contrary to the clearly expressed views of the people that federal presence should be reduced. With the severe restriction in available resources it does not seem logical to spend the public purse on unnecessary regulatory initiatives. In our view these hearings have amply demonstrated all the fundamental flaws of H.R. 1662 and we urge you not to report this measure from the subcommittees.

Thank you for your kind attention.

Appendix A

MCAT REFERENCES

1. Givner, N., and Hynes, K. New and Old MCAT: A Preliminary Comparison. Evaluation & The Health Professions., 1:95-100, 1978.

2. Pohlman, M., Richardson, M., and Pohlman, J. Use of a Simulated MCAT to Predict Real MCAT Scores. Evaluation & The Health Professions. 2:357-364, 1979.

3. Molidor, J.B., and Elstein, A.S. A Factor Analytic Study of the Old and New MCAT Examinations. Paper presented at the 18th Annual Conference on Research in Medical Education, Washington, D.C., October, 1979.

4. Deighton, B.C., Smith, D.U., and Gallagher, R.E. The Relationship Between Scores on the New Medical College Admission Test and Performance in the First Year of Medical School: A Preliminary Study. Paper presented at the 18th Annual Conference on Research in Medical Education, Washington, D.C., October, 1979.

5. Hynes, K., and Givner, N. An Empirical Investigation of Changes in MCAT Scores Upon Retest. Journal of Medical Education. 55:201-202.

6. Molidor, J.B., Elstein, A.S., and Scheifley, V. The Old and New MCAT Examinations: What Do They Measure? Paper presented at the American Educational Research Association Meeting, Boston, MA, April, 1980.

7. Cullen, T.J., Dohner, C.S., Peckham, P.D., Samson, W.E., and Schwarz, M.R. The Relationship of Scores on the New Medical College Admission Test to First Quarter Examination Scores. Journal of Medical Education. 55:393-398, 1980.

8. Friedman, C.P., and Bakewell, W.E. Incremental Validity of the New MCAT. Journal of Medical Education. 55:399-404, 1980.

9. McGuire, F.L. The New MCAT and Medical Student Performance. *Journal of Medical Education.* 55:405-408, 1980.

10. Jones, R., and Beran,R. Test Repeater Performance on the New MCAT. *MCAT Interpretive Studies Series* #80-1 AAMC: Washington, D.C., 1980.

11. Friedman,C., and Porter, C. *Incremental Validity: the Old and New MCAT's Compared.*. Paper presented at the 19th Annual Conference on Research in Medical Education, Washington, D.C., October, 1980.

12. Cullen,T.,Dohner,C., Peckham, P., and Samson,W. *Restriction of Range and the Predictive Validity of the New Medical College Admission Test.* Paper presented at the 19th Annual Conference on Research in Medical Education, Washington, D.C., October, 1980.

13. Cookson,F.B., Neil S.M., and Gupta, D.M. *An Analysis of the Predictive Accuracy of Various Criteria Used for the Selection of Medical Students.* Paper presented at the Annual Meeting of The Association of Canadian Medical Colleges, Winnipeg, Canada, October, 1980.

14. Harasym, P., and Fisher, L.A. *A Comparative Study of the Utility of the New and Old MCAT Scores as Predictors of Achievement: A Preliminary Study.* Paper presented at the 1980 Annual Meeting of The Association of Canadian Medical Colleges, Winnipeg, Canada, October 1980.

15. Essex,D., Sorlie,W.E., and Thompson, D.S. New MCAT Versus Old MCAT: Predictive Validity. *Journal of Medical Education.* 55:967, 1980.

16. Slotnick, H.B. Identifying and Remediating Specific Deficiencies of Entering Medical Students. *Journal of Medical Education.* 56:91-102, 1981.

17. Hart, M., Payne,D., and Lewis, L. Prediction of Basic Science Learning Outcomes with Cognitive Style and Traditional Admissions Criteria. *Journal of Medical Education.* 56:137-139, 1981.

18. Jones, R., and Thomae-Forgues, M. A Factor Comparison of Old and New MCAT Scales. *Journal of Medical Education.* 56:161-166, 1981 (also *MCAT Interpretive Studies Series* #81-1.)

19. Hynes, K., and Givner, N. Restriction of Range Effects on the New MCAT's Predictive Validity. *Journal of Medical Education.* 56:352-353, 1981.

20. Brooks, C.M.,Jackson, J., Hoffman H., and Hand, G. Validity of the New MCAT for Predicting GPA and NBME Part I Examination Performance. *Journal of Medical Education.* 56:767-768, 1981.

21. Veloski, J., Hojat, M, and Zeleznik, C. *The Overall Validity of the New MCAT.* Paper presented at the 20th Annual Conference on Research in Medical Education, Washington, D.C., October, 1981.

22. Hull,A., Calhoun, J., and Maxim, B. *Predicting Medical School Performance Using the Old and New MCAT.* Paper presented at the 20th Annual Conference on Research in Medical Education, Washington, D.C., October, 1981.

23. Beaton, J., Brooks, C.M., Maisiak,R.,Willcutt, H., and Linton, P. *Using the New MCAT and NBME Subject Examinations to Predict Performance in a Medical School Behavioral Sciences Course.* Paper presented at the 20th Annual Conference on Research in Medical Education, Washington,D.C., October, 1981.

Prepared by:
Division of Educational Measurement and Research
Association of American Medical Colleges
October, 1981

Mr. **Weiss**. Thank you very much Dr. Cooper. We appreciate you taking time away from the national meeting and coming to testify before our subcomittees. I think I'll reverse the process. Mr. Williams.

Mr. **Cooper**. With your permission I, of course, don't have the advantage of the hearing record here with me. I did call and have someone leave the national meeting and go to our offices and try and review the hearing records because I wondered if I'd had leave of my senses in making the statements that Congressman Williams said I had. We could not find it in the hearing records And I can assure you that we do have the lowest attrition rate in any professional school in the medical school. That is due not to the MCAT tests alone. It is due to the very detailed and extensive work that admissions committees do in evaluating the grade—the past performance, the MCAT examination, the activities of the students, the interviews, and letters of recommendations. And the MCAT examination in some of these preliminary tests which have been preliminary examinations which have been reported do indicate that it does add to the selection, but in no way do I think I have ever said that the MCAT was responsible for the lowering in the attrition rate. I didn't mean to say it anyway.

Mr. **Weiss**. You and Mr. Williams will explore that further as you both deem appropriate. Mr. Williams.

Mr. **Williams**. Thank you, Mr. Chairman. Mr. Cooper, my earlier reference was to a person I said at the time, I couldn't call his name or what association he was with and referred to the fact that he was representing some element within the medical community. I've since found the hearing record and for the record let me describe in more detail what took place.

Mr. James Graham, assistant secretary, Council on Dental Education of the American Dental Association appeared before the subcommittee in August 1979. And spoke at length about his association's support for admission testing. And indicated that the attrition rate was perhaps three times as great before the institution of a testing program then after.

As I recall he stated that their attrition rate was 20 to 25 percent before the dental admissions testing program was inaugurated and now was—or then was down to approximately a 7-percent rate.

My point to him was this, the association obviously had kept good records through the years about the numbers of students who dropped out of dental schools and I asked them if they had kept equally good records on why the students were dropping out or who the students were that were dropping out and his answer was—"No, it has not." I then went on to explore—now, let me comment about that. I think that that says a great deal in itself. I think that says reams about a lack of a pursuit concerning who's dropping out and why and then that lack of pursuit, in turn, says a great deal about the lack of an effort to prevent that attritition through some means other than simply not allowing that student into that school. Perhaps you'd care to comment at this point.

Mr. **Cooper**. I am sorry Mr. Williams you'll have to address those questions and comments to the dental schools because that does not apply to the medical schools. We do know how many drop out. We do know the names of the students. As a matter of fact we have extensive records going back many years with regard to the reasons for the dropouts, because some medical students interrupt their program to do a year of research. We know whether it's on the basis of academic performance, on the basis of family problems, and with the reducing support of medical students financial support by the Federal Government we're finding more students' having more difficulty in going through a regular series of the regular course of study. They have to get out and work. But we do have information. We do know how many drop out and for what reasons. There are some that are emotional reasons. And we have all of that. So, the statement that you made and attributed to the medical school does not apply to the medical schools.

Mr. Williams. Well, let me correct you again. I didn't attribute it to medical schools, I attributed it to a gentleman who represented an association of some element of the medical community and I was accurate with regard to that statement.

I appreciate the fact that your association takes every opportunity to maximize cpportunity for minorities and disadvantaged Americans. I'm from Montana. Our disadvantaged Americans are the original Americans. Very little is done by anyone to maximize their opportunity in medical schools or any other professional schools in this country. Very little is done for those people. It's not—there is no intentional grounding of the American Indian, it's just that—I think, as we'd all agree, enough isn't—isn't being done and many of us live with the American Indian or near the reservations or in States where the American Indians are very sensitive to the fact that that's a group—along with some others—Hispanics and blacks and others who really aren't being brought into the professional mainstream the way we might have done it in the past half century or century.

Mr. Cooper. Mr. Williams I couldn't agree with you more. I am a native New Mexican and I think we have more Indians than you do and I am very sympathetic and have worked with the Indians and the university which has done a great deal in New Mexico. And New Mexico, as you know, the majority of people in New Mexico are of Spanish origin. And I have an office of minority affairs. We have active programs in recruitment not only at the medical school but interesting these people in undergraduate preparation for medicine and as a selection of medicine as a career. I can say that Andy Young address—we—I just—yesterday afternoon and we have a minority program every year. And in our annual meeting to discuss what ways people have found to increase the interest of these students. How they have been able to retain them. Let me tell you again, sir, that the reduction in Federal support has played an unprecedented role in reducing their interest because the minority students simply: One, do not have the funding and second, they are not accustomed to the level of loans and with the interest rates of the loan which are 3½ points above the Treasury 91-day bill. They just can't consider that kind of funding. We have fought very hard to keep that exceptional need scholorship program in and it is completely inadequate.

That's how we can help get these people in medical schools and keep them there. In addition to that, we have developed a similar minority application—minority exercise for admissions committees.

Those—we have taken actual case records and showed these and these have been given without cost in most of the medical schools now in the United States and we're repeating them. We work with admissions committee members to show them that you have to use nontraditional criteria in admitting these students. And I'm very proud of the fact that in medical schools we have students with very high MCAT records and very high GPA records that are not admitted to any medical school because they are found lacking in the other aspects of the admission process. And we have whites with low MCAT's and low GPA's and we have minority students with low MCAT's and low GPA's admitted.

The admission committees of these countries in the medical schools work very hard and have not been slavish to the use of GPA's or MCAT's or any other single criteria as has been claimed, at least, in the discussions we had in New York that law schools do. We have not done that. And I think the evidence shows that we don't. I am as concerned and I am bilingual in Spanish. I learned to speak Spanish before I did English. I am as concerned as you are about the inadequacy of it. But Andy Young stood up in our minority meeting and said I'm not going to lay it on you. I was just out at Stanford in the graduation exercises for the MBA class and the number of blacks in that was really enormous. He said you're in a hell of a competition because we used to be able to only increase our social status through law or through medicine or through religion. Now, we have many other opportunities. And you're going to have a tougher and tougher time increasing the number of students interested in medicine. We are working on it though, very hard.

Mr. **Williams**. I applaud your efforts. Nay I ask you a question about one statement in your testimony. When you were referring to equalizing access to tests, you said that the legislation before us would exacerbate the problem and you used as a point of demonstration the fact that in New York only a small number are requesting test questions and answers. Fewer than 5 percent. And said that disproportionally more of those are from the higher socioeconomic groups and from those already achieving high scores on the test. That's saying that equalizing access will not benefit the disadvantaged.

I guess that raises two questions. First, do we deny access simply because the wrong people are taking advantage of it and second isn't this further demonstration that disadvantaged students—economically or socioeconomically disadvantaged students are because of the—because of their history less likely to vigorously pursue these tests? Isn't that further proof of the chilling effect that tests have on those who are already some frightened of testing authority?

Mr. **Cooper**. A responsible test agency would never use those same questions on subsequent tests. The only advantage of having the test before you, would be to show what questions you didn't want to study. That may be an advantage. But you would not repeat those questions.

We provide in an MCAT manual that has been submitted for the record, to Mr. Weiss, and to members of the committee and I'll be happy to submit it again, extensive information to the students. I think it's $4. That gives how the test questions were derived. What are the criteria and we do have very narrow criteria. They have to be questions that a student taking a first year course in chemistry not at Harvard, but across the country would have expected to have had knowledge about and be able to answer.

Second, it has it, relevance, as we have seen through the groups that the questions to what they think they should know to get into medical school. We tell them that. We tell them what areas we're going to examine in. We tell them the types of questions and show them examples of how those questions are developed and how they go about thinking about those questions. And then we give them an actual test. That test is the same test that they would have gotten by having one revealed except that it was one that we developed and used experimentally at the beginning which is exactly like the test we're using now. That test is as good as getting the one that you have. We will not use those questions either. But it gives you all of the information that's valid and useful from a test that is revealed.

By the way we have provided long before legislation was introduced in the State or here, we have provided opportunities for hand scoring of tests when students ask about them, say there are problems. We have provided that long before legislation is introduced. We have no problem complying with most of the interests which we agree with, that you had. We agree with them because we have instituted, we think, measures to think it. The only one that we disagree with is the main one. Now, there may be some others, but the main one is revealing the test.

We do comply with California. We have no problem in the financial information it's all available out there, but in revealing the test questions we can not do and maintain the kind of tests that we develop.

Mr. Williams. Thank you, sir, I appreciate the—appreciate the dialog here this afternoon.

Mr. Weiss. Dr. Cooper, picking up right there I must tell you that you have me confused, and that may not be your problem, it may be my problem. If I heard you correctly, you've just said that no responsible testing group would ever repeat the same questions.

Dr. Cooper. If it was revealed.

Mr. Weiss. Oh, OK.

Dr. Cooper. You have a responsibility to the other test takers, you know, that no one has an advantage.

Mr. Weiss. Then I did mis-hear you.

Dr. Cooper. If you reveal the questions you certainly would not incorporate them. We are re-cycling questions.

Mr. Weiss. OK. Now, what percentage?

Dr. Cooper. We have to.

Mr. Weiss. What percentage of your questions do you re-cycle?

Dr. Cooper. That, I'm sorry. We can provide it. It's just beginning because what we have done is over the years developed a series of test forms. We are now reach—and we would like to continue to develop new test forms. Now, we are running into the problems I have identified of the availability of additional questions which fit our criteria. I can't tell you the percentage, it is small so far, but I think will increase with time.

Now, I'm sorry that I don't have any of my staff here because I left them up there.

Mr. **Weiss**. Right. OK. Again to just review matters that you've testified to previously. The test areas that you advise your test takers that you'll be testing them on are, at least in the heart sciences: Chemistry, physics, biology, in the first year of the average college course; is that right?

Dr. **Cooper**. Well, the first course. It may not be in the first year.

Mr. **Weiss**. First course, right. OK. And what you've testified to is that the reason that you think that it would be inappropriate, wrong to disclose those questions and the tests is because there is a limited number of questions that you can ask within those subject areas, is that correct?

Dr. **Cooper**. It would be impossible for us and to maintain the tests. Let me say that the old MCAT tests which we felt was inadequate, the questions there we could have revealed. No problem because those questions can be developed in huge numbers. We were dissatisfied with that approach. That's why we spent $1½ million to change the test. The present questions are very restricted to, one—the criteria that you just said, and number two, to the examination by groups of medical faculty and so on that these kinds of questions are indeed, they believe, relevant to what a student should know in coming to medical school.

You know, what are the dynamics of the Space Shuttle. Those are good physics questions. They are not related to medicine. That's why the new test is the one in which we have the trouble. If we had not changed our MCAT test and had the same kind of general question, we would have less problem and, I don't know, we may have had no problem in revealing the question because you can generate those questions by the thousands.

Mr. **Weiss**. So, what you're saying is that you limit the area which is tested within the particular hard sciences?

Dr. **Cooper**. Yes.

Mr. **Weiss**. Such a narrow focus—such a narrow scope that in fact you have just a very limited number of questions that can possibly be asked?

Dr. **Cooper**. They certainly are much more limited than they were before. There's one other problem.

Mr. **Weiss**. But your answer is yes?

Dr. **Cooper**. Yes. One other problem and that is we have a more elaborate evaluation system of the questions, I think, than most other test groups have. We're not a testing agency. This is a small part of the overall activity.

Mr. **Weiss**. Right.

Dr. **Cooper**. Let me just go through what happens. We have tests written by faculty members. We have editors, AIR does under contract. They're reviewed by them. Then they're reviewed by members of a test committee. Then there's a field try out administration on a selective basis to examining populations. We get a group of college students that mimic——

Mr. **Weiss**. Doctor, I don't want to shortcut you but are you familiar at all with the process that the Educational Testing Service uses?

Dr. **Cooper**. Yes, sir. I'm not personally, no. Our people are.

Mr. WEISS. Your people are. Because I must tell you as one who was part of a congressional delegation of this committee that went down to Princeton and went through the process with them, I was impressed. They do exactly the kind of thing that you're talking about. They take people in the field, the topmost experts in the area, and they toss the questions around. They field test, they use them on an experimental basis. If they don't meet the criteria that they set up, they toss them out because they feel that in some way the way the question was phrased, perhaps, is wrong. Yet in spite of all of that they've had the recent experience within this single calendar year of having two questions disclosed by high school students who took in one instance the SAT and the other the PSAT that having gone through all the experts—hard scientific or mathematical questions——

Dr. COOPER. This doesn't go only through experts, Mr. Weiss. We have students take the test and the students tell us whether they are——

Mr. WEISS. Students too.

Dr. COOPER. Well, I don't know what—

Mr. WEISS. You're not suggesting that the students are likely to be more expert than the——

Dr. COOPER. No, but if there's ambiguity in the tests, students are bright. And if there's ambiguity in the question that makes the question change. If the student says you can't tell which is the right answer here for the following reason, that question is thrown out. That was the problem in the test. There was no correct answer.

Mr. WEISS. Are you saying—because again I—as I listen to your testimony are you saying that what our legislation seeks to achieve and what New York has done is appropriate for the other disciplines in other areas but that your test is so unique that it should not be applied to your area? Is that what you're telling us?

Dr. COOPER. We don't know about what happens in the other areas. That's for them to answer. We know that this legislation we can not live with.

Mr. WEISS. OK. Now let me see if I can ask you just a few other questions. Most of the colleges—most of the professional schools are going into an area of scarcity as far as student supply is concerned. I assume that that's not true as far as medical schools are concerned?

Dr. COOPER. No, sir we have had a traditional ratio of about two applicants per position. We are concerned, as I said before, about two things. One about applicants—in the applicant pool. And we have every applicant that applies we have biographical data on them so we can study the changes. We are concerned about the applicant pool coming from more affluent families. The family average income—average family income of those applicants is going up. And they are not—they are coming more and more—we have about a 13 to 15 percent physicians' children go to medical school. That's been constant for years.

But now other professionals—students—children of other professionals are going up and the manufacture, you know, the upper level and the ones from clerks, from agriculture, and from the lower middle class are not going. We are concerned about that and are going to work on it. One of the problems for that is the financial support. It's very difficult now for a middle-class family—particularly lower middle class that has accumulated debts hrough undergraduate college and then additional debts that they will accumulate through medical school.

We have students with $75,000 to $80,000 of debt before they leave medical school. So we are concerned about the nature of the applicant pool and we work on that and we are terribly disturbed and we have had two task forces. We led the way in 1971 and appointed a task force to say what is it we can do to get the under represented minorities in. We didn't achieve our goals and in 1976 we appointed another task force to say why didn't we achieve our goals. And now we're working on further input. We have an extensive program.

We are not happy with that component but as far as number it remains about 2 to 1. OK.

Mr. Weiss. Would you—do you have information as to what percentage of your applicants, in fact, attend the preparatory courses, the coaching schools?

Dr. Cooper. As I said, we ask a question on the application to the MCAT application, you know, the form that a student fills out. We ask them to tell us if they did. We have no other way to get that information.

Mr. Weiss. What do they tell you? What percentage do you get back?

Dr. Cooper. I can't give you the percent. I can get that for you and send it to you. But we did compare, as I pointed out in my testimony—oh, we took those that took the coaching course and took the MCAT more than once after the coaching course.

Mr. Weiss. Right. Right.

Dr. Cooper. And we studied that. But you know, that's very unreliable information. In spite of the fact we try and make sure that they understand we're not going to tell anybody about that coaching but——

Mr. Weiss. Right.

Dr. Cooper. But we can not, we do not have access to Mr. Kaplans', you know, list of students and so forth.

Mr. Weiss. Wouldn't you assume that the student who is competing to get into a more restricted area—medical school and who comes from the higher economic strata is more likely to attend—to take the coaching course than the student who comes to be admitted to college where in fact the colleges are looking for customers at this point rather than the other way around? And where there's a larger percentage of low-income applicants?

Dr. Cooper. I can speculate with you on it. I assume the more money you have, you know, the easier it would be. But our study shows it doesn't do much for you.

Mr. Weiss. No, no but I'm trying to establish first as to whether in fact, there is a high percentage or if you know whether there is a high percentage of the people who take your test who in fact attend the coaching school. And you're telling me that there is no way of knowing although you assume that if they got money they're more likely to attend those courses?

Dr. Cooper. On the other hand, you might expect that the whole educational process of those from more affluent families would be considered to be more adequate. And they might feel less need to have a coaching course than those who come from a disadvantaged educational background. Whether those latter can afford the course, I don't know. But, you know, we're just speculating here, Mr. Weiss, but maybe there would be fewer Harvard graduates—fewer Harvard graduates of whatever economic level take the course, than from East Branch State Teachers College take the course, if they could afford it. I think the Harvard students feel that they may be better prepared. You know, I don't know. We're just speculating.

Mr. Weiss. Well, I can only give you my experience with the cram courses for taking the bar. And my experience was that everybody took the cram course no matter what law school they went to. No matter how good their grades were. They wanted that extra little bit of insurance and if they could afford, they took it.

Dr. Cooper. I don't like to comment on one of our sister professions but it is—we do not have, you know, the cram course idea in law school—some of my lawyer friends that have gone through it say the law school you go to for 3 years and then you learn what you have to pass on the exam in the cram course. It has no relationship to what we learned in law school. I don't know what the situation is, but that is a tradition, I think, with law schools and law students.

Mr. Weiss. You are in a profession where youngsters graduating from college know that the scarcity of seats is such that they may be lucky if they get admitted to a foreign school, never mind to an American medical college.

Dr. Cooper. I think the latter is not so difficult.

Mr. Weiss. Well, given that kind of scarcity it would not surprise me at all if you had percentages running into the 1980's or 1990's of medical college applicants who took the cram course.

Dr. Cooper. I would seriously doubt that, Mr. Weiss. Seriously doubt it.

Mr. Weiss. Maybe we'll get some insight into that tomorrow. Let me touch on another area with you. You've indicated that in a sense this—and you correct me if I'm misstating or missummarizing your testimony. You've indicated that the legislation in a sense is a waste of time and you cite the ETS figure that only 5 percent of the people who took the test requested the answers. What you didn't cite was the 60-percent figure from the law school. Now, if you were able—if you were to make available to your medical college test takers the test upon requests or upon even the payment of a nominal fee to cover the expenses, would you speculate as to what percentage of the people who take the MCAT would in fact request their answers and the tests sheet, questions and answer back?

Dr. Cooper. I have no speculation about that, Mr. Weiss.

Mr. Weiss. Would you not guess that it would be extremely high?

Dr. Cooper. I really could not speculate. I don't know.

Mr. Weiss. Would you not like to find out?

Dr. Cooper. We can't find out because it's impossible for us.

Mr. Weiss. Well, you could run—you could run a test.

Dr. **Cooper**. We do have the MCAT manual which has a test in it.

Mr. **Weiss**. No, I know, I know.

Dr. **Cooper**. And which is no different from the test they would receive.

Mr. **Weiss**. You could run a test. You could run this kind of test: You could put on top of the test packet a sheet which said if we were to tell you that we would make available to you for $5 to $7, $10 the test—the scoring sheet with the keys and the questions and answers, would you request—would you be willing to pay for—would you be willing to request it? Why can't you run a test like that?

Dr. **Cooper**. The reason is that we, no matter what the answer was and you may be right or I may be right in saying I don't know, we could not ever fulfill that and maintain the MCAT exam.

Mr. **Weiss**. Well, the reason I asked you is because it seems to me that's unfair of you to come in here and say that the response to the SAT of ETS of 5 percent proves that on top of all the other problems with this legislation, it's unnecessary. You don't know whether in fact, in your area, it's unnecessary or not and I'm willing to bet you that you would get in excess of 50 percent. How is that? Probably closer to 95 percent, but in excess——

Dr. **Cooper**. We could both speculate, but you know——

Mr. **Weiss**. But you know, it seems to me that you could meet my challange very easily. You don't have to——

Dr. **Cooper**. I don't think it would be fair to the students, Mr. Weiss, because we could never fulfill——

Mr. **Weiss**. You would tell them—you could tell them listen we can't give it to you but supposing we made it available, would you be interested in ordering it. Be right up front with them.

Dr. **Cooper**. I frankly don't see the value of that kind of——

Mr. **Weiss**. Well, it would allow you to come in here after you got your response back with something a little bit more than you are drawing conclusions which don't apply to your area.

Dr. **Cooper**. I don't think that one answer would change our opinion. Whether it was 100 percent or 5 percent. The fundamentals of this legislation with regard to this specific test are simply that we cannot—we cannot comply and maintain the same test.

Mr. **Weiss**. Well, you know, the one thing I don't want to get into—with you is an argument that says yes it is and you say no it isn't. And so I'm not going to get into a discussion with you other than to comment in passing that your view of the finiteness of factual information to be tested is so strange coming from an educator—just is so strange. I mean it seems to me that people with imagination and with knowledge and with love of the subject area that they're testing ought to be able to find reams and reams of questions. You know, when I think of material that has been disclosed to the lay public over the course of this past year on public television alone, it seems to me that there are tremendous areas that you could test.

Dr. **Cooper**. Is that included in average 4-year courses in chemistry, physics and biology in the colleges?

Mr. WEISS. It's like saying well it's not in my catacism and therefore—I mean I don't know how to argue with that. It's a very difficult argument to pick up on.

Dr. COOPER. The other problem we have is we've tried to point out before, the number of tests we give and I must again repeat the elaborate validity and equating we use with—we had to put sample—we had to put questions in our tests which are not graded for the test but graded to evaluate one of the questions and No. 2 to equate with different tests. We have a limited—unless we make up and make the students sit for 2 days for an MCAT test, we simply with the other test components we have, cannot incorporate enough of these questions to permit us to validate and equate tests. So, it is not only the development of the tests, it's the elaborate scheme we have used to evaluate—validate and equate tests.

Mr. WEISS. How long did you—for how many years did you use the tests that was displaced by the current test?

Dr. COOPER. Dr. Moss, who by the way was a member of the faculty here at George Washington University first gave a test, it was called the Moss test. And then Dr. Moss came to the association—this was I must say before my time, in the 1930's. And we began giving the Moss test then, the medical college admissions test and that test went until we—it was modified. But there has been a lot of change in the science and so on of measurement. And we were concerned and began—it took us about 4 years to complete our studies and to get the input from the broadest group of people before we changed our test. So, we started in the early 1970's. So, it was between—let's say 1940 and 1970.

Mr. WEISS. Thirty years.

Dr. COOPER. Thirty years, yes, sir.

Mr. WEISS. And I bet you thought it was a pretty good test while you were giving it.

Dr. COOPER. Yes, but must tell you that there has been really significant improvement in our knowledge of evaluation and testing and ways that the knowledge has improved in all areas and we have taken that into account.

Mr. WEISS. The reason I asked the question is here you are now thinking that this is a pretty good test and I suspect that in about 10 years you'll come along and say we ought to throw that test out because it—we now have a really good test. You know, there are a lot of people in the psychometric field, I mean people who have devoted their lives to it, who come with some very serious and profound questions about the whole realm of testing. And I'm not sure that in fact it's accurate to say that we know so much more and therefore the test we're giving today are so much better. Indeed, I think you can spell out an argument for saying that there is so much doubt about the whole field of psychometrics that people who spent their careers have expressed grave doubts as to whether in fact there is inherent work or value of not—if more harm than good is not achieved by those tests. You know that I'm sure because you're in the field.

Dr. COOPER. You mean you're advocating abolition of standardized testing?

Mr. WEISS. I'm not because I'm not at that point at this stage, but what I'm saying to you is that I think it's—when you commit yourself, as you have, body and soul—heart and soul if you will to this specific test that you're giving currently, you may be letting yourself in for—for a big downer when, in fact, you realize that maybe it isn't what it's cracked up to be.

Dr. **Cooper**. We are anxiously awaiting the information that is coming from these independent studies outside on the value of this test in making decisions. That's for them to decide. Not for us. And we are awaiting it and we want to look at it along with you and anybody else with regard to what we've done. The preliminary indication is that the new MCAT is substantially better than the old MCAT for their purposes. Now, whether it is or not we'll have to wait the kind of studies that are being done.

Mr. **Weiss**. Why aren't you fellows pulling that suit together so that you have a deposition in New York as to what your rights are?

Dr. **Cooper**. I would never try and take the place of our legal firm. And I—that is for them to work out.

Mr. **Weiss**. That's the best answer you've given all day. Does council for the minority have any questions? No. Dr. Cooper——

Mr. **Williams**. Mr. Chairman will you yield?

Mr. **Weiss**. Of course, Mr. Williams.

Mr. **Williams**. Dr. Cooper, I found particularly instructive your concerns and the concerns of your association about the lack of appropriate funding for—for financial assistance to college students particularly minority and disadvantaged. And if it's appropriate, sir, I think that the chairman of the Subcommittee on Postsecondary Education, Congressman Simon of Illinois, and the chairman of the Subcommittee on Elementary, Secondary, and Vocational Education, Chairman Perkins of Kentucky would also find your association's specific thoughts in support of appropriate financial aid very instructive. And if you and your association think it would be appropriate, I'd encourage you to send a letter to those two chairmen and I'd appreciate having a copy of that letter outlining your concern about the lack of financial assistance in this area.

Dr. **Cooper**. Well, let me say, sir, we welcome that recommendation, of course. This legislative area is largely within Mr. Waxman's subcommittee in the House and Mr. Waxman is very sympathetic. He, as you know, has some difficulties in these kinds of programs now in the House and the Senate is very difficult. But our association this year took a stand—a step which had never been taken before and they identified two priorities that they would work on for Federal legislation. One was student assistance. Two was biomedical research. The reason for that is we feel that only the Federal Government can support. We did not work on capitation, on special project grants, on all of these other things. But we are deeply concerned about the lack of support for the medical student—the applicants and let me say we have—we have worked very hard in every committee except this one. But we would be very pleased to bring it to the attention of whoever can be helpful to us because we are deeply concerned.

Mr. **Williams**. Thank you. Thank you, Mr. Chairman.

Mr. **Weiss**. Again, Dr. Cooper we appreciate your taking the time away from your other activities to appear before us to testify.

Dr. **Cooper**. Thank you, Mr. Weiss.

Mr. WEISS. The hearing stands adjourned until 9:30 tomorrow morning. Same room.

[Whereupon the hearing was adjourned until 9:30 a.m. Thursday, November 5, 1981.]

[Additional information submitted for the record follows:]

PREPARED STATEMENT OF THE AMERICAN DENTAL ASSOCIATION

The American Dental Association opposes the enactment of the bills, H.R. 1662 and H.R. 1312, pending before the Committee. In order to understand the Association's objections to this legislation it is necessary to recognize the purpose and unique nature of the two national testing programs administered by the American Dental Association.

DENTAL ADMISSION TESTING PROGRAM

The Dental Admission Testing Program (DAT) is an entrance examination required by all of the 60 dental schools in the United States. The DAT was initiated in 1946 for the express purpose of reducing the unacceptably high attrition rate which was then occurring in dental education. It was not and is not intended as an instrument for predicting career success. During the 1930's and 1940's the annual attrition rate in dental schools approached 25 percent. Following the introduction of the DAT and its acceptance as an entrance requirement, the attrition rate was reduced to only 7 percent; a level which has generally been maintained since 1950.

Much of this success in reducing attrition results from the ability of the DAT to measure fine manual dexterity. In order to graduate from dental school, the student must have the proficiency to practice what is commonly called General Dentistry. General Dentistry, like the medical specialty of surgery, as opposed to the practice of most other professions, requires a high degree of manual technique proficiency. Not everyone has the fine manual dexterity to master the technical skills to practice dentistry. The DAT has developed non-verbal, perceptual and space relation tests which over the years have demonstrated a proven ability to measure and evaluate the applicant's fine manual dexterity. The format of these tests is non-verbal in nature. It is a series of drawings of two and three dimensional objects which the test subject must relate to. Because there is a finite limit to the number of ways to represent a box or triangle on a drawing, the number of test forms that can be produced is of necessity severely limited. If the Dental Admission Testing Program was required to disclose the actual tests after each administration, as the legislation now proposes, within a very short time these kind of manual dexterity tests would have to be discontinued. This issue is central to our objections to the legislation as presently structured.

The Dental Admission Testing Program is the only entrance examination that contains tests which measure abilities other than academic. A diminution in quality of the other portions of the DAT, for example its science tests or its reading comprehension tests, could be compensated for in part by separate criteria such as grades in college. Without the manual dexterity component of the DAT, however, dental schools would have no criteria to discriminate the ability to perform fine manual dexterity and as a consequence it can be predicted that the attrition rate in dental schools would again rise to unacceptably high levels. This reason alone compels the Association to oppose the legislation.

The Association further believes that the numerous disclosure requirements contained in H.R. 1662 are not relevant to the purposes of the bill. For example, Section 7(2) mandates disclosure of information regarding refunds given, fee waivers or reductions granted, income received, and expenses incurred by the test agency. The expenses are to be categorized as fixed or attributed to a specific purpose or program. It is unclear how this information will assist the Secretary of Education in furthering the purpose of the bill.

Some provisions of H.R. 1662 would be impossible to carry out. Section 3(a)(3) would require the Dental Admission Test Agency to determine the correlation between DAT scores and success in dentistry. There is, however, no practical way to measure success in a career, as each practitioner defines success according to his or her personal and professional goals.

In addition to the above noted concerns, the Association wishes to emphasize that the DAT program has for years provided a broad range of useful information to dental applicants.

1) It has released an example copy of an exact form of the Dental Admission Test which it has entitled "Dental Admission Test Preparation Materials." These materials are distributed free of charge to all applicants who register for the DAT. It has done this to neutralize the error measurement that has been introduced into the test scores by test preparation courses. Not all potential candidates, for reasons of economic circumstance or geographic location, have identical access to preparation courses. The Association feels that when a candidate studies for a test, he or she ought to be able to study the material which is actually contained in the test at the difficulty level it will be tested. This is especially important for the non-verbal, perceptual and space relation tests contained in the program. The American Dental Association also believes that it is assisting the dental schools to more accurately judge the ability of the candidates by eliminating the above types of bias that might be present in their admission decisions.

2) Annually, the American Dental Association conducts a complete statistical analysis of its tests along with validity studies which are participated in by all dental schools in the United States. The Statistical analysis contains a complete item by item analysis of the test to determine not only how reliable the tests were as

a whole, but also how each specific item in the test performed. The analysis also contains additional studies to determine the effect of repeating the test more than one time. These studies have been carried on since the inception of the program and are fully published and distributed free of charge to anyone who has requested copies of them. These studies along with other useful admission information are bound together under the title Handbook for Pre-Dental Advisors and annually distributed free of charge to all pre-dental advisors at all under-graduate, post-secondary institutions across the country. In this way, the American Dental Association feels that it fully informs and makes available all information that is necessary for an applicant or advisor to judge the quality or content of its' Dental Admission Testing Program. The additional, and we believe unnecessary, requirements of H.R. 1662 and H.R. 1312 would substantially increase the DAT application fee.

The Association also objects to the provision of H.R. 1312 which stipulates that knowledge examinations based for entry into an educational program or a career must be scored without relation to the distribution of scores produced by the examinations. Both the DAT Science exam and the National Board Examinations (noted below) are scored using a distribution-referenced scoring system. This scoring system allows for consistency among scores and equivalency of scores from various editions of the examination. Until criterion-based scoring procedures are refined to provide identical interpretation of scores from testing date to testing date, a distribution-referenced scoring system is the fairest and most reliable way of measuring the cognitive skills necessary to practice dentistry.

NATIONAL BOARD EXAMINATIONS

The Council on National Board Examinations was established in 1928 as a standing committee of the American Dental Association for the purpose of providing and conducting written examinations for use at the discretion of state boards of dentistry in licensing dental hygienists. Currently, over 12,000 candidates for dental licenses and almost 6,000 candidates for dental hygiene licenses participate in National Board programs each year.

The dental examination battery consists of seven, 100 item examinations, while the dental hygiene examination is a single examination consisting of 350 items. In the interest of serving the public, state boards of dentistry, candidates for licensure and the profession, these examinations are administered three times a year at testing centers throughout the United States. To insure examinations of high quality that are fair and just to participants, two to three new editions of each dental examination and one to two editions of the dental hygiene examination are developed each year. During any year, at least four different dental examinations are released with answers and statistical analysis of test items to licensure candidates, dental school faculty members and interested individuals. Dental hygiene examinations are also released regularly, but with less frequency due to the length of the examinations. The American Dental Association therefore believes that, through its current policy of releasing National Board examinations, it fully responds to the extent of the disclosure clause. Our concern relates to having undue restrictions placed on its examination program that would jeopardize the quality of examinations. Publishing all test items after their use would seriously deplete the test item files and cause examinations to become less reliable and less valid.

THE EDUCATIONAL TESTING ACT OF 1981

THURSDAY, NOVEMBER 5, 1981

House of Representatives, Subcommittee on Elementary, Secondary, and Vocational Education Jointly With the Subcommittee on Postsecondary Education, Committee on Education and Labor,
Washington, D.C.

The subcommittee met, pursuant to adjournment, at 9:30 a.m., in room 2175, Rayburn House Office Building, Hon. Paul Simon (chairman of the Subcommittee on Postsecondary Education) presiding.

Members present: Representatives Perkins, Simon, Miller, Weiss, Kildee, Erdahl, Petri, and DeNardis.

Staff present: John F. Jennings, counsel; Nancy L. Kober, legislative specialist; and William Blakey, counsel.

Mr. Simon. The Subcommittee on Elementary, Secondary, and Vocational Education and the Subcommittee on Postsecondary Education will resume its joint hearings on H.R. 1662, the Educational Testing Act of 1981.

I will enter this statement in the record which touches on what happened yesterday. Let me just say that our colleague from New York, Mr. Weiss, who has introduced the legislation has, I think, through the oversight of the subcommittee and through the action that's been taken by various organizations, caused considerable progress to be made in this area and those of us who are on the subcommittees and on the full committee appreciate that.

[The opening statement of Congressman Simon follows:]

Opening Statement of Hon. Paul Simon, a Representative in Congress From the State of Illinois, November 5, 1981

The subcommittee on Elementary, Secondary, and Vocational Education and the Subcommittee on Postsecondary Education will resume its joint hearings on H.R. 1662, the Educational Testing Act of 1981.

Yesterday's witnesses presented testimony on: One, why H.R. 1662 should or should not be enacted; two, the new policies enacted by the testing industry in response to New York State's test disclosure legislation; and three, the areas that have been amended in New York State's test disclosure legislation.

A number of yesterday's witnesses also talked about the area of coaching, an area on which many of the witnesses here today will present testimony. A number of the witnesses on yeaterday's panel indicated that they were not opposed to coaching. Dr. Cooper of the Association of American Medical Colleges testified that the Association found that through coaching the "higher gains were for content specific sections" of the MCAT (areas such as biology, chemistry, etc.) as opposed to the skills analysis section (reading and quantitative tests). Mr. Anrig from the ETS indicated that they do not tell students not to attend coaching schools or programs, but that he does question how much a student can improve in the short run in skill areas that should have been developed in the 4 years of high school study.

The role and impact of coaching is an area in which this committee seeks more information and therefore, we are happy to have before us today a group of people who are involved in the areas of testing and coaching. We have David A. Goslin who is the executive director, assembly of behavioral and social sciences at the National Academy of Science and his colleague at the Academy, Mrs. Alexandra Wigdor, study director, Committee on Ability Testing. We also have present Mrs. Lois Pines, former regional director of the Federal Trade Commission Boston regional office. The FTC's Boston office was responsible for release of a study entitled, "Coaching for Standardized Admission Tests." Our fourth witness will be Mr. Stanly Kaplan, founder of the Stanley Kaplan Educational Center. Commissioner James Miller of the Federal Trade Commission was unable to be with us today, but will submit written testimony for the record.

The witness I'm going to take first is Mrs. Lois Pines, the former regional director for the Federal Trade Commission from the Boston regional office. The Boston office was responsible for the release of the study entitled "Coaching for Standardized Admission Tests" and we're very pleased to have Mrs. Pines here.

I might add that we have a caucus right now of the Democratic members of the Budget Committee. We're putting together the second budget resolution and I'm going to have to excuse myself to Mrs. Pines and the other witnesses, at least temporarily. I'm going to turn the gavel over to Mr. Erdahl at this point. If you can come on over here.

We thank you for being here and my apologies.

Mrs. PINES. Thank you, Congressman Simon.

Mr. ERDAHL. Thank you, Mr. Chairman. We're delighted to have you with us, Mrs. Pines. You may proceed in any way that you wish. Your statement will be entered in its entirety into the record and you can give it or summarize it, whichever way that you might feel more comfortable.

Please proceed.

STATEMENT OF LOIS PINES, FORMER BOSTON REGIONAL DIRECTOR, FEDERAL TRADE COMMISSION

Mrs. PINES. Thank you, Mr. Chairman.

My name is Lois Pines and I'm delighted to have the opportunity to be here this morning to testify on H.R. 1662, the Educational Testing Act of 1981.

I firmly believe that the disclosures required by this legislation are essential to insure that the individuals who take standardized admission tests, as well as the institutions which utilize them, fully understand the benefits and the limitations of those particular tests. In my view, the issue really is a matter of marketplace fairness.

The 2½ years during which I served as the regional director of the New England office of the Federal Trade Commission made clear to me that without full and fair disclosure, consumers of any product or service cannot make a reasonable decision with regard to their purchasing, nor can they fully exercise their rights in the marketplace. In this particular case, where the service purchased can and very often does permanently affect the purchaser's educational and professional life, full disclosure is all the more important. Surely if consumers have the right to be protected by laws requiring disclosures in such areas as credit and product labeling, they have a right to full disclosure regarding the testing services that shape their futures.

Disclosure legislation in the education sphere is not new and as a State legislator from the Commonwealth of Massachusetts, I sponsored and wrote legislation that provided for the disclosure of students' cumulative educational record in the public schools, not only to students but also to their parents. That legislation became effective in Massachusetts in 1973.

Experts in the testing field who have previously testified before your committee, have cataloged the many dangers that are inherent in the widespread use of standardized admission tests. The predictive validity of the tests has been questioned. Concerns about racial, economic and cultural bias have been raised. Other critics have stressed that these multiple-choice tests cannot identify the truly creative or independent thinker; that such individuals may, in fact, be penalized by the tests. Personal qualities of importance, such as motivation, judgment, creativity can simply not be measured by these tests. My testimony today will focus on still another area of concern—the degree to which these tests are subject to special preparation or what is commonly referred to as coaching.

During my tenure at the Federal Trade Commission, my staff at the Boston regional office conducted an investigation regarding coaching for standardized admission tests. The investigation was spawned by concern that commercial testing schools were making unfair and deceptive claims of effectiveness.

The focus shifted, however, after the Federal Trade Commission staff conducted a statistical analysis which found that coaching at one commerical school was effective, raising its students' scholastic aptitude test scores an average of approximately 25 points in both the mathematical and verbal sections of the test. Staff reviewed other studies on coaching for the SAT and learned that other researchers had also found coaching to be effective. Our staff then undertook a review of the materials on the SAT distributed by the Educational Testing Service and College Board. The question became ETS and College Board fully and fairly disclosed the benefits of coaching for the SAT?

Our investigation culminated with the Commission's release of the Boston regional office's staff report on the Federal Trade Commission Investigation of Coaching for Standardized Admission Tests, copies of which you have. I believe that the staff report makes the need for H.R. 1662 absolutely clear.

One only needs to read the report's comparison of what the studies on coaching showed and what ETS and College Board were saying to conclude that students and educators have been misled. The report documents that the information sent to students and educators regarding coaching has been repeatedly in conflict with the research on that particular issue.

For example, in 1968 when the College Board trustees were telling students and educators that coaching would not work, that such endeavors corrupted education, researchers had already conducted studies showing coaching gains of 29 on math and 60 points on the verbal section of the test. After yet more research attesting to the effectiveness of coaching, research at times sponsored by College Board, students were still being told in the fall of 1979 that cramming does not raise scores.

Now ETS and College Board try to play a semantic game in their defense. They explain that all they meant by saying that cramming would not help was that a few hours of drill on sample questions would not help. I fear that the students and educators reading the ETS and College Board materials may not have been able to grasp such a fine distinction. When you have no other source of information, it's difficult to read between the lines. Coaching for the SAT can help. Coaching can mean significant score improvements, yet students and educators were led to believe that it would not help.

Full and fair disclosure has clearly not been the standard by which ETS and College Board have judged their material on the SAT. It is true that very recently, in the face of much public discussion and pressure, students have been given a more comprehensive statement on the coaching issue. However, absent a legislative mandate of H.R. 1662 insisting on full and fair disclosure as the norm, there will be no assurance that even these modest recent improvements will continue.

In fact, there is no reason to believe that misleading statements will continue to appear. ETS's reponse to the Boston regional office report is a case in point. A senior vice president for research at ETS was said to have argued that the FTC's findings were flawed, that ETS studies have indicated much lower gains, 10 points or so. Yet ETS continues to rely on studies which its own researchers have criticized as being irrelevant to today's test takers or to an assessment of the effectiveness of a substantive and well constructed coaching program. Moreover, a quote from one of ETS's public relations officials is even more astounding to me. On May 18, 1981, that official is quoted as saying, "We don't believe that short-term cramming can make a difference. These are aptitude tests designed to measure potential ability and can't really be studied for."

Mandating disclosure about the effectiveness of coaching would result in both practical and psychological benefits in the SAT context. On the practical side, students would fully understand the potential of coaching and be better able to decide whether such preparation is right for them. Also, schools would be better able to counsel their students on the question of preparation and decide whether to offer such preparation themselves. On the psychological side, students would be encouraged not to view their three-digit SAT scores as the final work of their mental capabilities. For if scores can be improved by coaching, the test is not measuring some unchanging and innate capacity to learn. It is not measuring aptitude.

The Boston regional office's staff report documents yet another reason why H.R. 1662 is essential. The report highlights the fact that a number of serious educational policy questions are raised by the finding that coaching for a standardized admission test, such as the SAT, can be effective.

For example, the implications of coaching for test validity must be examined. Questions are raised about the meaning of an examination designed to measure verbal and mathematical abilities said to be developed over a lifetime of learning if scores can be meaningfully changed in a few weeks or months of coaching. Validity may be further questioned if coaching is helping students gain inflated scores rather than helping them demonstrate their actual skills and abilities.

Perhaps the most important question raised in the report is the question of fairness. Is it fair to rely on a test that is subject to coaching, if that coaching is not equally available to all?

The evidence reviewed in the staff report indicates that the unequal access issue is important. All test takers do not have access to coaching. Commercial programs are available to students whose parents can afford to spend $200, $300, sometimes $400; but what about students whose parents do not have these resources? Schools in wealthy neighborhoods may be able to offer effective coaching programs, but financially pressed schools may not. Moreover, the report indicates that the very students most in need of coaching, those who may well have the most to gain by it, may have the least access to it. For example, there is some evidence that the degree to which one benefits from coaching is dependent on one's parental income and race.

Before the answer to the unequal access question can be found, however, some very fundamental questions about the test itself must be answered. What skills and knowledge does the SAT actually measure? What is the value of developing such skills and knowledge? And, finally, what is the value of training students to demonstrate those skills on any particular examination?

The staff report does not answer these important questions. Instead, it encourages experts in the testing field to devote their time and resources to finding these answers. In that regard, the Federal Trade Commission specifically asked the U.S. Department of Education to consider the important educational questions raised in the report.

At this time, however, it is difficult for independent experts to consider and conduct the necessary study and research. The test agencies exercise enormous control over what research is done and by whom. They have even controlled the extent to which research findings receive repeated citation and which findings are treated as no more than footnotes.

H.R. 1662 would break this monopoly power by assuring that independent researchers have access to the information they need to be able to study and evaluate the tests. Specifically, this legislation, H.R. 1662 would require the public disclosure of the test forms and the scored correct answers. It would require the full disclosure of all studies in which the test agencies have been involved, whether those studies support or refute their positions. If such information is not placed in the public domain the questions raised by our report cannot be pursued. Likewise, without this disclosure, the concerns raised by other critics, such as the possibility of economic and racial bias in the tests, cannot be effectively addressed. The impact of standardized admission tests is far too strong to permit only those with a self-perpetuating interest in the outcome to review the validity of the tests.

For the role that standardized test scores play in the educational and professional lives of millions of Americans should not be under-rated. Here, too, the Boston regional office's staff report provides, I believe, very important information. The report notes that a 1979 survey conducted in part by College Board revealed that over half the competitive colleges responding viewed test scores as a very important factor, and almost 2 percent considered test scores the single most important factor in making admissions decisions. Approximately 40 percent of both the public and private institutions responding to the survey indicated that they had minimum SAT scores below which applicants generally are not considered eligible for admission.

TTS and College Board respond to critics' concern about the importance of the SAT by explaining that many colleges are simply no longer that selective. Such statements are of little comfort to the student with a combined score of 975 who wishes to attend a school that has a minimum requirement of 1,000. It is little comfort to students whose SAT scores are not high enough to assure admission to the college of their choice because they were unable to obtain effective coaching. These test takers have a right to review the questions asked and to assure themselves that their responses were correctly scored. They have a right to learn from their past mistakes and by the experiences of others. Section 3 of H.R. 1662 would give test takers, the students, those rights.

Millions of students and parents are traumatized each year in confronting the SAT. Yet the marketplace cannot now guarantee that the best testing instruments are being used to evaluate our children. For students cannot choose to ignore these tests and their advocates cannot now effectively review them. Only if all participants in the testing marketplace have access to the test questions and the answers, and to all relevant research, can we be sure that the tests used are valid and fair. A dollar figure can simply not be placed on the lost potential, if access to college and to the professions is controlled by inadequate measurement tools.

In conclusion, I believe that the Boston regional office's report represents an important question, an important addition to the debate on standardized admission tests. I hope that the report and my remarks will be helpful to the committee in its deliberations on H.R. 1662.

I'm personally very disappointed that the Federal Trade Commission and Joan Garrity who was the lead attorney for my staff, who prepared this report are not with you today and I very much hope that the committee will have the opportunity to hear from the Federal Trade Commission with regard to this matter.

I'd be delighted to answer any questions and very much hope that members of this particular committee will believe that the work done by the Boston office of the Federal Trade Commission was valuable and that you'll assist in insuring that the Boston regional office is one of the 10 regional offices that the Federal Trade Commission continues to have in place working to protect the public.

[The prepared statement of Lois Pines follows:]

PREPARED STATEMENT OF LOIS G. PINES, FORMER DIRECTOR, BOSTON REGIONAL OFFICE, FEDERAL TRADE COMMISSION

Good morning Mr. Chairmen and members of the Subcommittees. My name is Lois G. Pines. I am happy to respond to the Subcommittees' invitation to speak today on H.R. 1662, the Educational Testing Act of 1981.

I firmly believe that the disclosures required by H.R. 1662 are essential to insure that the individuals who take standardized admission tests, as well as the institutions which utilize them, fully understand the benefits *and* the limitations of those tests. In my view, it is a matter of simple marketplace fairness.

The two and one-half years during which I served as the Director of the Boston Regional Office of the Federal Trade Commission made clear to me that without full and fair disclosure, consumers of any product or service cannot make reasonable purchasing decisions nor can they fully exercise their rights in the marketplace. In this particular case, when the service purchased can, and very often does, permanently affect the purchaser's educational and professional life, full disclosure is all the more important. Surely if consumers have the right to be protected by laws requiring disclosure in such areas as credit and product labelling, they have a right to full disclosure regarding the testing services that shape their futures.

Disclosure legislation in the education sphere is not new. As a state legislator in Massachusetts, I sponsored legislation that provided for the disclosure of students' cumulative public school records to students and their parents. That legislation became effective in 1973.

Experts in the testing field who have previously testified before you have catalogued the many dangers inherent in the widespread use of standardized admission tests. The predictive validity of the tests has been questioned. Concerns about racial, economic and cultural bias have been raised. Other critics have stressed that these multiple-choice tests cannot identify the truly creative or independent thinker; that such individuals may, in fact, be penalized by the tests. Personal qualities of immense importance, such as motivation and judgment, can simply not be measured by these tests. My testimony today will focus on still another area of concern—the degree to which these tests are subject to special preparation or "coaching."

During my tensure with the FTC, my staff at the Boston Regional Office conducted an investigation regarding coaching for standardized admission tests. The investigation was spawned by concern that commercial coaching schools were making unfair and deceptive claims of effectiveness.

The focus shifted, however, after FTC staff conducted a statistical analysis which found that coaching at one commercial school was effective, raising its students' Scholastic Aptitude Test scores an average of approximately 25 points on both the mathematical and verbal sections of the test. Staff reviewed other studies on coaching for the SAT and learned that other researchers had also found coaching to be effective. Staff then undertook a review of the materials on the SAT distributed by the Educational Testing Service and College Board. The question became—have ETS and College Board fully and fairly disclosed the benefits of coaching for the SAT?

Our investigation culminated with the Commission's release of the Boston Regional Office's Staff Report on the Federal Trade Commission Investigation of Coaching for Standardized Admission Tests. I believe that the staff report makes the need for H.R. 1662 absolutely clear.

One only needs to read the report's comparison of what the studies on coaching showed and what ETS and College Board were saying to conclude that students and educators have been misled. The report documents that the information sent to students and educators regarding coaching has been repeatedly in conflict with the research on that issue.

For example, in 1968 when the College Board trustees were telling students and educators that coaching could not work, that such endeavors "corrupted" education, researchers had already conducted studies showing coaching gains of 29 points on math and 60 points on the verbal section of the test. After yet more research attesting to the effectiveness of coaching, research at times sponsored by College Board, students were still being told in the fall of 1979 that cramming does not raise scores.

Now ETS and College Board try to play a semantic game in their defense. They explain that all they meant by saying that "cramming" would not help was that a few hours of drill on sample questions would not help. I fear that the students and educators reading the ETS and College Board materials may not have been able to grasp such a fine distinction. When you have no other source of information it is difficult to read between the lines. Coaching for the SAT can help—coaching can mean significant score improvements—yet students and educators were led to believe that it would not help.

Full and fair disclosure has clearly not been the standard by which ETS and College Board have judged their materials on the SAT. It is true that very recently, in the face of much public discussion and pressure, students have been given a more comprehensive statement on the coaching issue. However, absent the legislative mandate of H.R. 1662 insisting on full and fair disclosure as the norm, there will be no assurance that even these modest recent improvements will continue.

In fact, there is reason to believe that misleading statements will continue to appear. ETS's response to the Boston Regional Office report is a case in point. A senior vice president for research at ETS was said to have argued that the FTC's findings were flawed, that ETS studies had indicated much lower gains—10 points or so. Yet ETS has relied on one study for which it says it cannot even find the underlying data. Its own researchers have criticized many of the other studies on which it relies as being irrelevant to today's test takers or to an assessment of the effectiveness of a substantive and well constructed coaching program. Moreover, a quote from one of ETS's public relations officials is even more astounding. On May 18, 1981, that official is quoted as saying—"We don't believe that short-term cramming can make a difference. These are aptitude tests designed to measure potential ability and can't really be studied for."

Mandating disclosure about the effectiveness of coaching would result in both practical and psychological benefits in the SAT context. On the practical side, students would fully understand the potential of coaching and be better able to decide whether such special preparation is right for them. Also, schools would be better able to counsel their students on the question of preparation and to decide whether to offer such preparation themselves. On the psychological side, students would be encouraged not to view their three digit SAT scores as the final word on their mental capabilities. For if scores can be improved by coaching, the test is not measuring some unchanging and innate capacity to learn. It is not measuring "aptitude."

The Boston Regional Office's staff report documents yet another reason why H.R. 1662 is essential. The report highlights the fact that a number of serious educational policy questions are raised by a finding that coaching for a standardized admission test, such as the SAT, can be effective.

For example, the implications of coaching for test validity must be examined. Questions are raised about the meaning of an examination designed to measure verbal and mathematical abilities said to be developed over a lifetime of learning if scores can be meaningfully changed in a few weeks or months of coaching. Validity may be further questioned if coaching is helping students gain inflated scores rather than helping them demonstrate their actual skills and abilities.

Perhaps the most important question raised in the report is the question of fairness. Is it fair to rely on a test that is subject to coaching, if that coaching is not equally available to all?

The evidence reviewed in the staff report indicates that the unequal access issue is important. All test takers do not have access to coaching. Commercial programs are available to students whose parents can afford the $200-$300 prices, but what about students whose parents do not have such resources? Schools in wealthy neighborhoods may be able to offer effective coaching programs, but financially pressed schools may not. Moreover, the report indicates that the very students most in need of coaching—those who may well have the most to gain by it—may have the least access to it. For example, there is some evidence that the degree to which one benefits from coaching is dependent on one's parental income and race.

Before the answer to the unequal access question can be found, however, some very fundamental questions about the test itself must be answered. What skills and knowledge does the SAT actually measure? What is the value of developing such skills and knowledge? And, finally, what is the value of training students to demonstrate those skills on any particular examination?

The staff report does not provide answers to these important questions. Instead, in encourages experts in the testing field to devote their time and resources to finding the answers. In that regard, the Commission specifically asked that the Department of Education consider the important educational policy questions raised in the report.

At this time, however, it is difficult for independent experts to conduct the necessary study and research. The test agencies exercise enormous control over what research is done and by whom. They have even controlled the extent to which research findings receive repeated citation and which findings are treated as no more than footnotes.

H.R. 1662 would break this monopoly power by assuring that independent researchers have access to the information they need to be able to study and evaluate the tests. Specifically, H.R. 1662 would require the public disclosure of the test forms and scored correct answers. It would require the full disclosure of all studies in which the test agencies have been involved, whether those studies support or refute their positions. If such information is not placed in the public domain, the questions raised in the report cannot be pursued. Likewise, without this disclosure, the concerns raised by other critics, such as the possibility of economic and racial bias in the tests, cannot be effectively addressed. The impact of standardized admission tests is far too strong to permit only those with a self-perpetuating interest in the outcome to review the validity of the tests.

For the role that standardized admission test scores play in the educational and professional lives of millions of Americans should not be underrated. Here, too, the Boston Regional Office's staff report provides important information. The report notes that a 1979 survey conducted in part by College Board revealed that over half the competitive colleges responding viewed test scores as a very important factor, and almost two percent considered them the single most important factor in making admission decisions. Approximately forty percent of both the public and private institutions responding indicated that they had minimum SAT scores below which applicants generally are not considered eligible for admission.

ETS and College Board respond to critics' concern about the importance of the SAT by explaining that many colleges are simply no longer that selective. Such statements are little comfort to the student with a combined score of 975 who wishes to attend a school that has a minimum requirement of 1,000. It is little comfort to students whose SAT scores are not high enough to assure admission to the college of their choice because they were unable to obtain effective coaching. These test takers have a right to review the questions asked and to assure themselves that their responses were scored correctly. They have a right to learn from their past mistakes and the experience of others. H.R. 1662 would give test takers those rights.

Millions of students and parents are traumatized each year in confronting the SAT. Yet the marketplace cannot now guarantee that the best testing instruments are being used to evaluate our children. For students cannot choose to ignore these tests and their advocates cannot now effectively review them. Only if *all* participants in the testing marketplace have access to the test questions and answers, and to all relevant research, can we be sure the tests used are valid and fair. A dollar figure can simply not be placed on the lost potential, if access to college and to the professions is controlled by inadequate measurement tools.

In conclusion, I believe that the Boston Regional Office's report represents an important addition to the debate on standardized admission tests. I hope that the report and my remarks are helpful in your deliberations on H.R. 1662.

I strongly urge that you support the bill.

Mr. ERDAHL. Thank you very much for your statement, Mrs. Pines. The staff has just informed me that there is a reasonable possibility that the FTC will be submitting some written testimony.

You know it, I'm sure, that Mr. Weiss has arrived, the principal author of the bill and from the majority side and I would at this time relinquish the gavel to my colleague from New York.

Mr. WEISS. Mr. Erdahl, thank you very much for assuming the chair and I want to apologize for not having been here earlier. The Senate Committee on Government Operations is holding some hearings which affected some of my constituents and I had to testify so that's my reason for not being here.

I want to welcome you, Mrs. Pines, for being with us and for your participation in this issue over the course of the years. I don't know what to expect from the FTC these days, so I'm not sure that the word that they're going to make a submission is necessarily good news. We shall wait and see.

Mrs. PINES. I might remind you, Congressman Weiss, that when the report was issued by the Federal Trade Commission in May of this year, it was done so with a unanimous vote of all 5 members.

Mr. WEISS. Yes; but there have been some changes.

Mrs. PINES. I have noted that.

Mr. WEISS. Yes; in any event, we hope for the best.

Do you have information or would that really have been at staff level as to the manner in which the study was set up, can you testify to that at all as to who was interviewed and on what basis, and so on?

Mrs. PINES. When the original study was done by the Federal Trade Commission, I was not the Regional Director and I would prefer to submit to you, which I'd be happy to do, material with regard to the specifics of that particular study.

I initiated this study, which was based upon the original study that I believe your staff has with regard to a finding that coaching can be effective. That study was done as I arrived at the Federal Trade Commission, I didn't participate in setting that study up in terms of establishing the boundaries of that study.

Mr. **Weiss**. All right.

Was yours basically a statistical and interview study?

Mrs. **Pines**. No; no, our original study was a very carefully done statistical study that took a great deal of time.

Mr. **Weiss**. We had some information yesterday, I think that Commissioner Anrig gave us indicating ETS's conclusions as to the difference which coaching made. And again, I don't have the numbers in front of me. My recollection is that he said that it was a difference of 13 points on the verbal and 21 on math. Is that right or do you remember? I think that those—on the average, right.

The statistics that you have indicate a much larger variable than that.

Mrs. **Pines**. Unfortunately, I have not had the opportunity to read or hear Commissioner Anrig's comments. I understand that those figures were provided for 20-hour programs as opposed to longer programs. Did he deal with, for example, a 40-hour course?

Mr. **Weiss**. It is significant that ETS has been much more specific over the course of the past year or so up through yesterday as to exactly what they meant and what they mean by the kind of coaching which has no impact, little impact, greater impact. And I think that we're perhaps getting close to a point where we're using language in the same way and I think that partly has been the problem that we've had in arriving at mutual conclusions in that we were using—we seemingly were using the same language but in fact we were not. I think that we're getting there.

Mrs. **Pines**. I guess I would ask for a clarification, are they saying—are they agreeing that coaching—did they refute the Federal Trade Commission's report yesterday?

Mr. **Weiss**. No. No. They—we really I don't think had very much to do by way of refutation. I think what they're—what Commissioner Anrig was saying was that in a sense the discussion about coaching is sort of irrelevant and second that the amount of difference it makes obviously depends on how much coaching or preparation work is done and I don't think anybody has disputed that.

The earlier position that ETS had taken was quite different as the testimony indicates.

Mrs. **Pines**. Absolutely.

Mr. **Weiss**. OK. Mr. Erdahl.

Mr. **Erdahl**. Thank you very much, Mr. Chairman, I thank you again for being with us today.

From your testimony, maybe I have arrived at an improper conclusion, are you against testing per se or the standardized tests?

Mrs. **Pines**. Absolutely not.

Mr. **Erdahl**. What are some of the alternatives to testing for a school or institution to at least forecast for themselves and for the students potential academic success? You mentioned in your remarks, I believe that some schools establish a cutoff for admission of a combined SAT score of 1,000. The person that scored 975 was out. What of the person that scored 300?

Mrs. **Pines**. Our report did not deal with whether or not there should be testing, our report and my comment dealt with the issue of whether coaching could be effective in assisting a student in raising his or her scores. And once you determine or you accept the fact as I believe now has been done, that coaching can be effective in assisting a student in raising his or her scores. Then that, I believe, calls into question the validity of that particular test with regard to what it is that test is intended to do.

I am not here recommending that no testing or no standardized testing be utilized, but what we've attempted to do is to raise questions which we think educators are best able to deal with.

ETS for years has represented that coaching would not help and that all one needed to do was to have a good night's sleep and a good breakfast before you took that particular test that day and what the Federal Trade Commission looked into was the validity of that particular statement and we concluded in our report that they have misrepresented the evidence with regard to coaching.

Mr. **Erdahl**. Mr. Chairman, that leads to another question, when does one, either the taker or the giver, when do we cross the line from studying and cramming and coaching? I think that may be a fuzzy question, but I think it's fundamental. If it's wrong to be coached for a test, is it wrong to try to cram or to study?

You mentioned about the students traumatized by ETS tests. I suppose a lot of us when we were going to school were traumatized by any tests. I'm not sure that that is our goal to get away from that potential trauma or pressure. That maybe is good. Could you tell me when do we cross the line between studying, cramming and coaching?

Mrs. **Pines**. Well, as Chairman Weiss indicated, we're dealing with semantics in terms of what this issue is all about. ETS basically stated that for years that one could not prepare for their test. What we established was that indeed there have been various studies that appear to be quite valid that do substantiate that coaching has been helpful to students in increasing their scores.

Certainly I would imagine that most educators would encourage students to continue to study and to learn and to improve themselves educationally. And when one particular test has the magnitude that the SAT has, then obviously a great deal of scrutiny needs to be given to that test and the circumstances that surround the giving of that test.

It's not an ordinary test, because the SAT is the single most important factor used by 2 percent of the colleges that responded to a survey done by the College Board.

Mr. **Erdahl**. That statistic surprised me because I would have thought that it would have been used by at least 50 percent. Other criteria surely come in whether it's high school scores, evaluation by teachers and a lot of things. I'm not trying to challenge you, but I frankly was surprised that only 2 percent used that as a major, I assumed that it was used by more.

Mrs. **Pines**. The single most important factor however let me check my statistic here, but many of the schools used it, most of the schools use it and the question is to what degree do they rely upon it.

Mr. **Erdahl**. We had some hearings on this last year where many of the university schools and others testified they used high school rank and other indications but that testing is one of the tools used I supposed for them and for the student to give some indication of potential academic achievement and whether he or she is going to survive in that school.

Mrs. **Pines**. That's correct, however, when 56 percent of the school, it's almost 57 percent of the schools, cited the admission score as a very important factor and 33 percent considered them one of several factors. An enormous amount of weight is obviously being placed on one test. Oftentimes to the exclusion of the academic record of that student over their entire high school career.

And for that reason we felt that this was a very important issue that the public needed to have addressed and examined carefully.

Mr. **Erdahl**. I don't think it really excludes the other things, but maybe—and you're entitled to your opinion, maybe that's too much weight put on it.

You have mentioned the study several times. I have not read it yet, but I have a copy of it and evidently, as I understand it, this was released under a blue cover and designated as a staff memorandum. Its date was September 1978, but it was released in May of 1979. I believe that was the time you were the Regional Director up at Boston, is that correct?

Mrs. **Pines**. I had become Regional Director in March of 1979.

Mr. **Erdahl**. OK.

Then I'm going to read a disclaimer that appears on the cover of the report and would like to have you comment on that.

It says, note, this memorandum is being placed in the public domain because it might make a contribution to the public debate on the issues addressed. The views and conclusions expressed in this memorandum are those of staff members of the Boston Regional Office of the Federal Trade Commission. Those views and conclusions have not been adopted by the Commission or any individual commissioner.

Moreover, the Commission specifically believes that some of the conclusions in the study are not supported by the evidence obtained in the investigation.

Could you comment on that disclaimer and why didn't at least one of the members of the Federal Trade Commission express a willingness to be identified with the staff report?

Mrs. **Pines**. I don't believe that the issue was placed before the Commission with regard to a decision on the adoption of the report. The Federal Trade Commission prepares many staff reports on numerous or hundreds of issues over a period of time. Economic analysis as well as a statistical analysis, such as that one.

And the issue was not whether or not the Commission should adopt the report, it never was brought to the Commission for such adoption.

Mr. **Erdahl**. So, you claim that the disclaimer is not really a judgment on it but just the fact that they didn't consider it?

Mrs. **Pines**. Absolutely and indeed, this particular report which I submitted to you today was indeed adopted by the Commission and was voted to be distributed by a unanimous report and this report relies upon the statistical analysis that was done in the report which you have before you.

Mr. **Erdahl**. Again, I'm not trying to be contentious here, but staff has just handed me a copy of the brief report and it says the views contained in this document are those of the staff of the Federal Trade Commission, the views have not been formally adopted by the Commission. If I could respond to your other observation, somebody on the Commission or some other staff members must have made a judgment because the closing sentence is, moreover, the Commission specifically believes that some of the conclusions in the study are not supported by the evidence obtained in the investigation.

Is that disclaimer placed on by the Commission or a commissioner or other staff?

Mrs. **Pines**. That question ought to be asked of the Federal Trade Commission when they appear before you. I am not here representing the Commission, I am here as an individual this morning. I was the Director of the Office that prepared this report and not that report.

Mr. **Weiss**. By this report you mean the small staff summary, is that the idea?

Mrs. **Pines**. Correct.

Mr. **Erdahl**. I'm still a bit confused. Just so we have it straight because I understand that this was released in May 1979. The date on it is September 1978, but it was released at a time during the period that you were in fact—I'm not trying to cross-examine you, I don't want to make it appear that way, because you've been a very cooperative witness and you gave us some good testimony. But it was released during the time that you evidently were the Director but you said you were not involved in the first report?

Mrs. **Pines**. That's correct, I was not personally involved in the preparation of the original report. The report that I believe that you have before you was released by the Bureau of Consumer Protection in Washington.

Mr. **Erdahl**. I don't know about that, but, Mr. Chairman, it says it's a staff memorandum of the Boston Regional Office of the Federal Trade Commission, the effects of coaching on standardized admission examinations.

Mrs. **Pines**. That's correct. Let me try to explain.

Mr. **Erdahl**. OK.

Then I'm probably taking more of my time, but please make an explanation and I yield back to my colleagues, but go ahead.

Mrs. **Pines**. Part of my tenure at the FTC, the Boston Regional Office did an analysis. The Office of the Bureau of Consumer Protection in Washington did a reanalysis and they issued an additional report. At that time I had become the Regional Director of the Boston Office, however, I personally did not participate in any way with regard to the issuance of that report or the research done on which that report was based.

Mr. **Erdahl**. OK.

Maybe this will clarify it. Are the report or reports, are they the result of a group of studies, two studies or one study?

Mrs. **Pines**. The original statistical analysis was done once and it was reexamined and reanalyzed several times. This particular report with regard to coaching for standardized admissions tests was based—was an additional report that was based upon the findings in the statistical analysis that coaching could be helpful and effective for students.

It was also based upon additional research done by my staff with regard to the other studies that had been done by EPS and others on the issue of coaching per se. So, that I was not involved with the study done by the Federal Trade Commission initially prior to my tenure relative to the analysis of the effectiveness of coaching for a group of students that were carefully analyzed.

Mr. **Erdahl**. OK.

Thank you very much. Thank you, Mr. Chairman.

Mr. **Weiss**. Mr. DeNardis, I'm sorry, it's a long morning.

Mr. **DeNardis**. Mrs. Pines, thank you for your testimony; this is a subject of great interest to me as a former educator, teacher, social scientist for the past 16 years before my election to this body. I'm apt to want to scrutinize testimony on this subject, therefore, more carefully than might be normally expected. And I was interested in the exchange between you and Congressman Erdahl. I think there is one further question in that if I may. Immediately inside this blue front cover of the report there appears a full page statement from the Bureau of Consumer Protection, which I assume is in the FTC——

Mrs. **Pines**. In Washington.

Mr. **DeNardis**. The headquarters bureau?

Mrs. **Pines**. I don't happen to have that report with me because I was not planning to testify on that report.

Mr. **DeNardis**. This is dated May 1979. You were the Director at that time?

Mrs. **Pines**. I had just become the director. Although, as I told Congressman Erdahl, I was not involved in the preparation of that particular report in any way. That was issued out of Washington and although the initial research and work had been done by the Boston regional office, it was before I was personally there and I had no affiliation with that report.

Mr. **DeNardis**. Well, in the interests of truth in testifying, let us just clarify this and say for the record that there does appear a statement by the Bureau of Consumer Protection dated May 1979, which further disclaims the Boston office report and in fact, says there are several major flaws in the data analysis making the results unreliable. I think that should be noted for the record.

I think you have been very forthright, however, about your involvement with this and I'm not questioning that, I'm just questioning the document. You were the director for a time and during that time you reissued a report——

Mrs. **Pines**. It wasn't a reissuance of a report, it was a different analysis.

Mr. **DeNardis**. A different analysis.

Mrs. **Pines**. This is a different report than that report.

Mr. **DeNardis**. But as Congressman Erdahl has noted, the disclaimers still appear.

Mrs. **Pines**. That is common practice by the Federal Trade Commission.

Mr. **DeNardis**. At any rate, what we're trying to do is to establish your interest and involvement——

Mr. **Miller**. What is the common practice?

Mr. **Weiss**. Are you asking Mr. DeNardis to yield to you?

Mr. **Miller**. Yes; if you would just for a minute because I have the same confusion. We have a disclaimer on the cover——

Mr. **DeNardis**. I would be very glad to yield to my colleague from California.

Mr. **Miller**. We apparently have not cleared up the problem of what is the impact or the status of the report with the disclaimer on the cover. You mention that the report has been adopted unanimously. Is it a matter of changing the cover to reflect that? I think we have some confusion on the panel here still.

Mrs. **Pines**. Congressman Miller, the common practice of the Federal Trade Commission is to put this disclaimer on practically everything that comes out of the Commission. And the Commission did vote to issue this report by unanimous vote and in addition, the Chairman of the Commission pursuant to the recommendations of the staff wrote to the Department of Education referring this report. Copies of this report were sent by the Chairman of the Commission to the members of the Congress who are interested in this particular issue.

Mr. **Miller**. Well, for the purposes of this committee, if I can interrupt you, what does that mean, that they stand behind the report, that they have found no flaws in the report as they have analyzed it, and they think the methodology is correct, what does that mean? Does it mean that they basically believe it's correct but it's not going to get their good housekeeping seal of approval? I don't understand what that means.

Mrs. **Pines**. The Federal Trade Commission did vote unanimously, to forward this report to Members of Congress, to issue it to the public and to forward a letter to the——

Mr. **Miller**. But it hasn't been adopted by——

Mrs. **Pines**. Secretary of Education, but indeed to the best of my knowledge such reports are never dealt with any differently than this report.

Mr. **Miller**. I guess it's a luxury that those of us in politics don't understand. I don't understand how I could put something out of my office and say it doesn't reflect my views and it hasn't been adopted, but it's out there. Where's the accountability for the report, that's all I'm trying to establish.

Mr. **Weiss**. Well, if the gentleman will yield, I assume that if the FTC were to send a representative we could clarify that position, but we're not sure at this point as to whether they will in fact send someone in person.

Mr. **Miller**. Just send the committee a letter explaining the procedure and how this is done. I'm not trying to attack the validity of the report.

Mr. **Weiss**. We will ask, so that the record is complete, for the FTC to explain what their various notations and grading on reports which are released by their office.

[Information referred to follows:]

Hon. PAUL SIMON,
Chairman, Subcommittee on Postsecondary Education, Committee on Education and Labor, House of Representatives, Washington, D.C.

DEAR MR. CHAIRMAN: Thank you for your letter of November 10, 1981.

In response to your questions:

(1) FTC reports are published with a disclaimer whenever the reports have not been adopted by a vote of the Commission.

(2) Reports issued by regional offices are handled in the same way.

For example, the statement Mr. Muris, Director of the Bureau of Consumer Protection, submitted to your joint hearings contained a disclaimer for the very same reason: to distinguish staff views from formal Commission positions.

Please let me know if you have any further questions.

Sincerely yours,

JAMES C. MILLER III,
Chairman.

Mr. WESS. Mr. DeNardis.

Mr. DENARDIS. Yes. So, we have—we have established at least this, that one; you are an independent witness, you are appearing here on your own and second, you have presented us with some reports which have not been adopted by the Federal Trade Commission, but which may be interesting and illuminating within a narrow range of reliability.

Mr. WEISS. You may expand on that.

Mr. DENARDIS. And that third, I understand that you are now engaged in a political campaign in the State of Massachusetts. Is that true too?

Mrs. PINES. I'm planning to be.

Mr. DENARDIS. Good luck. I don't know what you're running for or what label you bear, but we're glad to welcome you to the political arena.

Mrs. PINES. I've been there before.

Mr. WEISS. Mrs. Pines, if you—I'm sure Mr. DeNardis would want you to have the opportunity to respond to the comment question that he just put to you in its entirety in anyway that you would like, right?

Mr. DENARDIS. Certainly, Mr. Chairman. I was going to go on but you may respond to that, but I was going to go on and ask you another question if I might with respect to your testimony on page 7.

Mr. WEISS. Before you do that, do you want to expand on the characterization that Mr. DeNardis made of what you testified to as far as these reports are concerned?

Mrs. PINES. I'm concerned with Congressman DeNardis's comments obviously. An enormous amount of work and effort went into this analysis, I oversaw the work personally and the Federal Trade Commission has adopted this policy with regard to issuing their disclaimers on everything from press releases to all of the staff work that's done on an ongoing basis and it's common ordinary practice for the Federal Trade Commission. I'm not commenting on the validity of it or the acceptability of it or the process and certainly I would ask the FTC, although they have no reason to respond to me at this time. To forward to you——

Mrs. WEISS. We'll do that.

Mr. PINES.They're common ordinary the analysis of this procedure that has been adopted.

But I would certainly hope that the members of this committee would not allow the adoption of a disclaimer to detract from the work that was so carefully done and analyzed and read carefully by all members of the Commission before the Commission took action based upon this particular matter.

Mr. DENARDIS. I would say that it's not merely the lack of adoption by the Commission that concerns us but it is their comment about the validity or lack thereof of the conclusions based on the methodology and the evidence examined. And since you have now just stated that you personally supervised the work, it is now appropriate for me to ask you what background and experience you have in the design of research and the analysis of empirical data, which I would not have asked had you not said you personally supervised the work.

Mrs. PINES. I'm sorry, Congressman, you obviously misunderstood what I attempted to say. I did not supervised the initial report, I am only referring to this particular report——

Mr. WEISS. Instead of—may I make a suggestion, instead of talking about this report and that report, perhaps if we put a date on the report. When was the small report that you have anything to do with issued?

Mrs. PINES. The May 1981 report.

Mr. WEISS. All right.

So, we're talking about the 1981 report, but you had nothing to do with the 1979 report, is that right?

Mrs. PINES. That is correct.

Mr. WEISS. Thank you.

Mr. DENARDIS. Thank you, Mr. Chairman, for helping us clarify this. On page 7 of your testimony, the second paragraph, you talk about the role of standardized admission—the role that standardized admission test scores play in the educational and professional lives of millions of Americans. You go on to make the point that over half of the competitive colleges responding viewed test scores as a very important factor and almost 2 percent considered them the single most important factor in making admission decisions.

The Weiss bill as I read it and as I understand it does not affect how admissions counselors and admissons committees deal with the standardized test scores and I just wanted——

Mrs. PINES. Absolutely.

Mr. DENARDIS. OK.

So you're not saying that, however much you might be raising that as a point and taking issue with this——

Mrs. PINES. No.

Mr. DENARDIS. You're not taking issue with it. Then let's simply clarify it, that this bill does nothing to correct the inappropriate use of test scores?

Mrs. PINES. I'm not suggesting that it's an inappropriate, I'm not making a value judgement with regard to how the educational institutions utilize the particular—any tests or other educational information that they utilize which to make their decisions. However, we are merely stating the conclusion of the ETS or the College Board study where schools were interviewed and asked to what degree they relied upon the SAT.

And based upon the responses of the institutions merely point out that the institutions are relying very heavily or many of them are relying very heavily on this particular test and thereby putting forth the question if they rely on the test and coaching can be helpful and students or some students or many students do not have access to coaching, what does it mean for the future of that particular student who may score at a lesser level than would be needed to gain admission to that particular institution.

Mr. DeNardis. Well, what would the adoption of this bill do to remedy that situation?

Mrs. Pines. I believe that the adoption of the legislation would provide for fuller and fair disclosure to the public with regard to the test results and test scores and in addition the information that the statistical information and other data gathered by the test agencies and thus give independent researchers the opportunity to do the analyses that might be helpful in terms of making the educational judgements or adding to the discussion with regard to the educational arena with regard to these educational policy issues.

I think that the bill is important because it provides students with an opportunity to see their tests, ascertain where they went wrong and in some cases as I'm sure you're well aware, when ETS chose to do that, in one instance, a particular youngster found that he had answered the question correctly although ETS had called it wrong, thus forcing ETS to alter or to raise 240,000 youngster's scores because they had answered the question in a similar vein.

Mr. DeNardis. Be that as it may, isn't it true now that there is a considerable amount of voluntary disclosure as the result of the pressure of this legislation and legislation in several States? The fact of the matter is there is considerable amount of voluntary disclosure at this point.

Mrs. Pines. There is voluntary disclosure. There is some voluntary disclosure at this point. I think, however, that it's very important that this legislation be passed so as to insure that limited voluntary disclosure is expanded to cover all tests in all circumstances.

ETS has for many years failed to disclose much of the information that I personally should have been made public. ETS chose not to report the findings of several studies with regard to coaching in their own literature. I think that it's important for the public to have full access to all the information so that an appropriate analysis can be made. It would end the monopoly that ETS presently has with regard to the information gathered by them and the analyses done by them and I think that in—that that's very important when we're talking about an issue that affects millions of students who take this test. And their futures are obviously critically dependent on what happens to them in the taking of this one test in their lifetime.

Mr. DeNardis. Thank you. I have additional questions but I would yield to the chairman who would like to pass on to another member.

Mr. **Weiss**. Thank you very much, I appreciate your cooperation, we'll come back to a second round, if any member wishes to question.

Mr. Miller.

Mr. **Miller**. Thank you, Mr. Chairman. You at some point in the report—I think it's on page 12, try to show the preciseness with which—within some institutions, the SAT is used. Either with the combined grade point score or at our own University of California, you have a sliding scale depending upon the combined grade point and SAT. Then in your executive summary you suggest that the College Board has submitted some descriptive materials prepared for the 1980-81 testing year, talking about whether a 10 point or 15 point gain could be made by various coaching programs.

Actually, for a combined 25 point increase and then a suggestion that the Commission feels in some instances that may even be higher and there's also a question raised in the executive summary whether that first suggestion of 10 or 15 points is in fact going to be increased.

I still have a concern that I don't see that it is answered either by the legislation or by the testing service and that is what happens to kids that can't partake in coaching. And as we've heard over the life of the hearings, this does have a serious impact on the future of these young people.

I encountered it last week or the week before when my son came home and told me he was going to take the PSAT, I guess, and then suggested what coaching could do for him. I tried to suggest that yes, no, maybe—a lot like the report here.

But he was convinced if he enrolled in a coaching course he was going to pick himself up 50 or 100 points. I assume that was based upon locker room talk at the high school, but it was obviously very prevalent that this is what everybody was going to do.

I'm very concerned to what happens: one, the extent to which students are being misled about what coaching will or will not do and to the extent to which those students who are not able to participate in coaching, if it does have an increase of somewhere between 25 points and 100, what happens to those students. And I echo that concern and I think we can get as bogged down as we want on disclaimers, but this has not yet been answered by this Educational Testing Service and it has not yet been answered by the Commission or this committee.

I dare say there is a whole host of students that are being left out.

Mrs. **Pines**. I would agree with you, Congressman Miller, and that's the reason that I chose to come today on my own because I felt that this was a very important issue.

There are many students who are unable to perform well on a standardized test. However, the Federal Trade Commission merely pointed out the evidence so that the educators and the Department of Education and the Congress too, could deal with this particular subject. The Federal Trade Commission and certainly myself are not educational experts and we did not put ourselves forward as educational experts, however based upon the information that we have, we believe that these are important educational issues that need to be dealt with and we hope that by putting forth this report we would encourage the educational people to consider this issue carefully.

Mr. MILLER. Well, I wish they would and I think to some extent the committee may be misled in that we have had testimony from certain law schools and medical schools and certain universities suggesting that what they really do is to look at the whole person and make these individualized determinations.

I suspect that for the vast majority of students who apply to colleges and universities that is not the case. And that in fact there is a formula worked out and you can engage in special pleading if you don't get in and try to present your case.

I think we're being misled when universities and law schools tell us that that's in fact what they really do, I just don't believe that's true for the overwhelming number. It sounds good and it kind of makes the issue go away because nobody is really being harmed. But I don't think that's so in your evidence that you presented out of the University of California undergraduate application packet of 1981-82 or the others suggest that that is not the case.

Mrs. PINES. That's is why we felt that it was important.

Mr. MILLER. I don't know how many thousands of students apply to the University of California system each year as freshman, but I'd be very surprised if each one of them got a little interview here to determine whether they really had inner qualities that would get them 2.79 to 3.01. I just don't believe that that happens and therefore, I think you have what appears as a discriminatory test. Where the discrimination can possibly be overcome by the expenditure of money, at that point I think we've got a policy problem that we've got to cope with.

I don't know if our testing bill is the right answer or the wrong answer, but I think we've got a problem here that suggests that the students are being excluded or potentially being excluded for that reason.

Mr. WEISS. Will the gentleman yield on that?

Mr. MILLER. Yes, I'd be delighted to yield, I just want to thank you very much for your testimony and your statements.

Mrs. PINES. Congressman Miller, I appreciate your comments very much. I too have a youngster who is now a senior in high school and waiting for the decision from the colleges to which he has submitted an application. And I share your concern as a parent who had gone through this process and you'll have another year to cogitate on it.

I'd also say that I had the opportunity to appear before you several years ago with regard to the asbestos issue and very much appreciated the leadership that you and Congressman Perkins provided in that area.

Mr. MILLER. Thank you for your nice comments, but you'd love to see what EPA has done to that one. Don't go into those old schools right at the moment.

Mrs. PINES. Well, I'm happy to say in Masachusetts where I was a State legislator at the time and the chairman on the Commission on Asbestos that we have done a great deal and we've analyzed all of our schools and we've made those changes where needed and we continue to monitor those schools that have potential problems for the future.

Mr. **Miller**. God bless you.

Mr. **Weiss**. Yes, I thank you for yielding. Just on the question of coaching and what the benefits of this legislation is to those who otherwise would not be able to afford, even if they were aware of the fact that coaching was available or that it could be helpful.

We had testimony at our hearings in July from a young man who is the executive director of a college and career counseling program in East Harlem, New York City and what he pointed out was that because of the disclosure provisions of the New York State law, he is now able to compete very effectively with Mr. Kaplan for example, that—without having the kids subjected to the fees, which they could not otherwise afford; even though Mr. Kaplan's program provides scholarships, he can't cover everybody.

And so indeed the disclosure provision provide an opportunity for kids who otherwise would not have had the opportunity for coaching and could not compete on an equal level.

Mr. **Miller**. Well that's very helpful. I guess many of us in politics draw on our own experiences. I remember when I was graduating from college and took the LSAT. My scores really stunk so I took them a second time on the theory that you could improve them. They went down and my father begged me, don't take it anymore, he said, or I won't be able to help you. So he helped me and I got a special interview with the dean of the law school. My father was chairman of the Senate Finance Committee, they needed a building and I got into law school. I did very well in law school, but I would have been excluded based upon that test. I didn't know there was coaching, I don't know if coaching would have helped, but there was a lot of suggesting that coaching wouldn't even have helped.

Later when I graduated from law school, there came the State bar exam. I signed up with one of these bar review courses with a fellow who nobody had ever heard of but gave guaranteed results and it worked. I passed the bar on the first time. And I suspect that the manner in which this fellow approached it for the bar would be considered outrageous by those people that gave the bar, by those people who now practice law.

I'm very concerned that people still could be precluded because I still have friends that have taken the bar 14 times and still trying to, jump the hurdle that is precluding them from practice, from professonal life, from entering professional school and now even being precluded from entering college and universities. And I'm very disturbed about that system.

I'm not suggesting that we're going to develop a system that treats everyone as equals because clearly that's not going to be the case. But I think we've got to ask ourselves when a single event in a person's life can change the course of that individual's like, you'd like to know at least that you give him the best shot you could. Whether that's providing Federal scholarships to Mr. Kaplan's school, I don't know. But you'd like to believe that they all started out with the same shot.

Thank you.

Mr. **Weiss**. Thank you, Mr. Miller. Mr. Petri.

Mrs. **Pines**. Could I—excuse me—could I just respond to the intitial question.

Mr. **Weiss**. You still remember it?

Mr. **Miller**. This is a test.

Mrs. **Pines**. I'd like to underline what you've indicated, Mr. Chairman, with regard to the importance of the legislation as it relates to coaching because the legislation in section 4 specifically requires that the statistical data be made available and the analysis be made available and thus there'd be an opportunity to better understand or at least have our educators better understand, statisticians better able to make judgments with regard to what kind of coaching does help. And that information is not presently available.

So I believe the bill is important in that it provides that disclosure and independent researchers would have that and analysis would be able to be made which presently cannot be made because the information is not available from the Educational Testing Service.

Mr. **Weiss**. Mr. Petri.

Mr. **Petri**. Thank you, I'd like to thank you for taking the time to come and offer testimony here today. I'm sitting here listening to all this trying to figure out whether there is a clear line between coaching and education? I mean we know that seniors do better than first graders and you could say that they've been coached for 12 years and then take the test. Kids go to summer schools that are oriented more towards the exams—is that something that should be a concern?

I went to school in the Midwest where only three of us in my class even heard of or even took the educational tests and then there are kids going to extra schools and such who seem to go year after year through educational courses oriented toward that. Somehow I suppose you could say that those of us from other parts of the country are at a disadvantage if we're not oriented in that way, but is that really true?

Is there some clear idea of what coaching is and what education is and how you separate one from the other?

Mrs. **Pines**. Well, I think that's really a major issue that we're dealing with. Coaching at the present time is available to very few youngsters nationwide, and it appears that those youngsters that have had access to coaching have benefited from it or many of them have benefited from it.

And it appears that some kind of coaching or drill may not be of help and in fact, in the analysis done originally by the FTC, there were two coaching programs that were studied. One which was helpful and one which was not helpful. So the issues really become what kind of coaching is helpful and to whom is it available. And are you precluded from access to that coaching because of your income or your rates, the way you live, geographic location.

And those are the issues that I personally believe need to be raised and that's the reason the report was published and the reason that I felt that it was important to come here to discuss it.

Mr. **Petri**. And so what we're trying to do is to address ourselves to someone who attempts to help a student get a better score apart and aside from whatever school or whatever other thing they happened to be engaged in and the person does that for money, is that the idea?

Mrs. **Pines**. Well, the issue is really what kind of coaching or education is available to students generally. Some public schools as well as private schools provide a coaching program as part of their academic curriculum in some parts of the country.

But again, on a limited basis.

Mr. **Petri**. But should we regulate that? What are we trying to get at here?

Mrs. **Pines**. What I'm trying to do is——

Mr. **Weiss**. Gentleman, if you will yield. Perhaps I can place into context what we're doing here today. We've had a series of hearings starting in I think July of 1979, on the issue of the truth—so called truth in testing legislation. And we touched on various aspects of the field and various aspects of the legislation.

One of the issues that has cropped up and which is in fact touched on the legislation as far as disclosure is concerned is on the effect of coaching on the taking of various tests, disclosing that information.

We have not had a session that was devoted specifically to the coaching issue although from time to time in some of the panels we've had some expiration of the issue and so what we did for today is to put together a forum for exploring that issue and that's what we're doing today. We're trying to cover various aspects and indeed we'll have some experts testifying in addition to Mrs. Pines, who will be touching on different aspects of the issue.

Thank you, Mr. Petri, Mr. Kildee.

Mr. **Kildee**. I have no questions, Mr. Chairman.

Mr. **Weiss**. Mr. Erdahl.

Mr. **Erdahl**. Thank you very much, Mr. Chairman, maybe I should plead a bit guilty, because we got off on to—maybe they're not tangents but a little different area about the report itself. I think, Mr. Chairman, you properly zeroed in on the focus of today's hearing which is the consideration of the impact of this coaching on the testing and I would share that concern. I'm not sure if your bill addresses it, but it seems to me that we have to be concerned that if specific coaching is available to students that might have the money to pay for it or might live in an area where it's available, then we have questions. How do we address that? Should that be part of the high school curriculum, should it be something mandated by the Congress, which we've kind of shied away from doing? Should it be subsidized in some way? I think it raises some very serious question.

Mr. Chairman, I'm not sure——

Mr. **Weiss**. If you'll yield just on that one point.

Mr. **Erdahl**. Of course I yield.

Mr. **Weiss**. Again, at page 4 of the legislation under section 3, sub 4, sub C, the requirement is for statement—disclosure on the registration form of statements concerning the effects on and uses of test scores including C, the extent to which test preparation courses improve test subject scores on average expressed as a percentage.

And again, all that this legislation does is to seek disclosure, we are not regulating anything.

Mr. **Erdahl**. Thank you very much. I understand that but it does raise a broader question if we find that in fact your bill if it's passed would come up with that conclusion. As Mr. Miller said, maybe as members of the Congress we can afford to send our sons and daughters to a coaching school maybe other people can't. I think that your——

Mr. WEISS. Next bill.

Mr. ERDAHL. Next bill. OK.

Mr. WEISS. Not this one.

Mr. ERDAHL. OK.

But then, Mr. Chairman, I would ask this as a request that I received from our colleague, Bill Goodling, who couldn't be here today that came through staff just a few minutes ago, We have had what could be described as some rather strong charges against ETS for maybe their failure to disclose and so forth. I understand we have a person in the audience who I have not met, Dr. Mesic I believe it's pronounced. I think that he is regarded as ETS's coaching expert and I was wondering if it would be proper at this time or something during the course of this hearing to see if he has any response to any questions.

I see from the witness list that this individual is not included on your formal list, Mr. Chairman.

Mr. WEISS. I think that ETS will probably tell you that if it wasn't for the honor, they'd just as soon walk. They have been given perhaps more of an opportunity to participate in the series of hearings than they would have liked if they had the chance to make the determination.

Obviously if the time permits and we've gone through the witnesses whom we specifically invited, we'd be pleased to put them on. We had, as you know, the president of ETS testifying yesterday.

Mr. ERDAHL. Yes, I was here.

Mr. WEISS. On some of those issues, but I do have people from the National Academy of Sciences who have been invited to come down and I want to put them on next. I'd like to put Mr. Kaplan on and then if after that we still have time and the members want to hear additional witnesses, I would be pleased to be as accommodating as you want.

Mr. ERDAHL. Thank you very much, Mr. Chairman, I think that's most reasonable and again, ma'am, thank you very much for being with us today.

Mr. WEISS. Mr. DeNardis.

Mr. DENARDIS. Following up on your review of section 3A, subsection 4, and subsection C, which deals with the provision on coaching. I guess I want to clarify with you first, that this has to do with the test agency. The test agency shall provide this information and make statements concerning the use of test scores, but wouldn't that, Mr. Chairman, be dependent upon the private coaching agency providing that information to the test agency?

Mr. WEISS. Again; if the gentleman would yield. Part of the problem that we've had and I think that we've developed it so that we've almost come to a meeting of the minds is that there had been information gathered by various researchers including ETS and its researchers as to what the impact of various kinds of preparatory courses would be.

However, in submitting the information to the registrant, eventual test taker, the impression until we started in this whole process, indeed up to a year or a year and a half ago that ETS created was and quite liberally was that coaching was in essence a future effort and we had a panel about a year and a half ago in which we explored the extent to which some of their own research had not been disclosed and so what we're really asking is that if you've got the information as to what the impact is, for crying out loud, don't con people into not taking preparatory courses or coaching courses when in fact you've got a test which may determine what you're going to be doing for the rest of your life.

And I think that on the basis of the exchange that we had yesterday and even before that with ETS, I think we've fairly well come along at this point where they recognize what they were talking about was short term less than 4 hour drills. What everybody else is talking about were in fact courses which ran 20, 40 hours or longer. And everybody agrees now that if you take the longer courses, yes, you're going to improve your scores. There may still be a difference as to exactly what the average may be but we're coming pretty close to that.

And all that we're asking is the disclosure not from the coaching schools, but from the objective researchers as to what the impact of those courses are.

Let the—let the test takers make their own judgment as to whether they want to take coaching courses or not. I don't want to force them. That information ought to be disclosed honestly and objectively.

Mr. DeNardis. As I recall from my son's packet last year as he prepared to take the test, isn't there a reference made to that in the packet that is distributed?

Mr. Weiss. We're getting there, yes. If you were to check the stuff they had 2 or 3 years ago.

Mr. DeNardis. Well, I think, Congressman Weiss, you have done the whole field of testing a great service. I really mean that. I think this bill—the mere existence of this bill has had a salutory effect on the whole testing field.

I think the question before the committee is should we actually enact it into law.

Mr. Weiss. You want to stop me on the 1-yard line. Is that what you're saying?

Mr. DeNardis. Well, very often the threat of legislation brings about reform without its actual enactment.

Mr. Weiss. Right. Thank you.

Mr. DeNardis. I have no further questions.

Mr. Weiss. Mrs. Pines, thank you very, very much, we appreciate your perseverance and you're taking your own time to come here and testify.

Our next witnesses are David A. Goslin, executive director, assembly of behavioral and social sciences, National Academy of Sciences and Alexandra Wigdor, study director, committee on ability testing, National Academy of Sciences.

I welcome you both and appreciate your patience in standing by.

STATEMENT OF DAVID A. GOSLIN, EXECUTIVE DIRECTOR, ASSEMBLY OF BEHAVIORAL AND SOCIAL SCIENCES, NATIONAL RESEARCH COUNCIL, NATIONAL ACADEMY OF SCIENCES, WASHINGTON, D.C.

Mr. GOSLIN. Thank you, Mr. Chairman, it's a pleasure to be here. My name is David Goslin, I'm the Executive Director of the Assembly of Behavioral and Social Sciences, which is one of the eight major divisions of the National Research Council, the principal operating arm of the National Academy of Engineering.

In accordance with the Academy's Congressional Charter enacted in 1863, the Assembly responds to requests from executive branch agencies and from the Congress for advice on specific scientific and technical questions involving behavioral and social sciences and on occasion takes the initiative in proposing studies on topics of national concern.

Currently, the Assembly is reponsible for supervising the activities for 30 major committees and panels, involving over 500 distinguished behavioral and social scientists and other experts, and addressing a wide range of significant topics including national statistical programs, child development research and public policy, law enforcement and the administration of justice, population and demography, basic research in the behavioral and social sciences, human factors, vision and hearing, substance abuse, and aging. As you know, members of Academy committees serve without compensation othe than travel and other necessary expenses.

The Assembly's Committee on Ability Testing was formed in 1978 to undertake a broad examination of the uses and potential abuses of standardized ability tests in American society. Support for the committee's work was provided by the Carnegie Corporation of New York, the National Institute of Education, the National Institute of Mental Health, the Office of Personnel Management and the Ittleson Foundation.

Given it's broad mandate, a special effort was made at the outset to appoint as members of the committee a diverse group of outstanding scientists and other experts, a majority of who had not been involved in the existing controversies about tests.

Mr. WEISS. We do have a live quarium at this point, I'm going to remain and will continue the hearing but anybody who wants to go, we'll be here when you get back.

Mr. GOSLIN. The chairman of the Committee on Ability Testing, Professor Wendell Garner, is a distinguished experimental psychologist, a member of the National Academy of Sciences and former dean of the graduate school at Yale University. His schedule, unfortunately, did not permit him to be present this morning.

Among other members of the committee are Burke Marchall, former Solicitor General of the United States and now professor of law at Yale University, Philip Kurland, distinguished professor of constitutional law at the University of Chicago, John Tukey, professor of statistics at Princeton that many think is the most distinguished statistician in the United States, Sociologist William Sewell of the University of Wisconsin and chairman of the recent National Commission on Research, Marcus Alexis, former Commissioner of the Interstate Commerce Commission and Historian

Oscar Handlin. Two distinguished psychometricians, Lee Cronback of Stanford University and Mel Novick of the University of Iowa also served on the committee. A complete list of committee members will be submitted as part of Alexadra Wigdor's subsequent testimony.

The committee has completed its report, which is now undergoing final editing prior to formal publication by the National Academy Press. In light of the report's significance and the complexity of the issues it addresses, the full report will be released to the public on February 2, 1982 at a day-long symposium sponsored by the National Academy of Sciences. I and the Academy's new President, Dr. Frank Press, hope that you and/or members of the committee staff will be able to attend. We will be in touch with your committee staff regarding specific plans for this event.

Mr. WEISS. I thank you for the invitation.

Alexandra Wigdor, study director for the committee is with me this morning to offer substantive tstimony on behalf of the committee. She's a legal historian, trained at the University of Maryland and has directed staff work for the committee during the last 1½ years.

Finally, let me add a word of personal perspective on this project and on the legislation being considered by the subcommittee. For the first half dozen years of my career as a sociologist, from 1961 to 1968, I realize that's 7 instead of 6, I participated in the series of studies conducted by the Russell Sage Foundation on the social consequences of standardized testing. In addition to authoring two of the several volumes that resulted from this work, I was responsibile for preparing the foundation's report, "Guidelines for the Collection, Maintenance and Dissemination of Pupil Records," which subsequently served as a basis for what is now known as the Buckley amendment. It was clear to us at the foundation in the early 1960's that the problems associated with the widespread use of standardized tests in our society transcended the many technical issues of validity and test bias that were being argued then and the have been so hotly debated in the intervening years. At the heart of the controversy that has surrounded testing are the central questions of what qualities, talents, and skills the society values most and how best to nurture them. As John Gardner put it in 1958, "the greatness of a nation may be manifested in many ways in its purpose, its courage, its moral responsibility, its cultural and scientific eminence, the tenor of its daily life. But ultimately the source of its greatness is in the individual who constitute the living substance of the Nation." If this is true; then the means that are used to differentiate one individual from another and to reward those attributes most highly valued will determine not only the fortunes of individuals, but also that of the society. It is my hope and expectation that the Academy's forthcoming report will make a significant contribution to what will and should be a continuing national debate on these issues.

With your permission, Mr. Chairman, I should now like to introduce Alexandra Wigdor, Study Director of the Committee on Ability Testing. Following her statement, we are prepared to respond to questions that members of the subcommittee may have.

Thank you.

[The prepared statement of David Goslin follows:]

Prepared Statement of David A. Goslin, Executive Director, Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, D.C.

I am David Goslin, Executive Director of the Assembly of Behavioral and Social Sciences, one of eight major divisions of the National Research Council, the principal operating agency of the National Academy of Sciences. In accordance with the Academy's Congressional Charter, enacted in 1863, the Assembly responds to requests from executive branch agencies and the Congress for advice on specific scientific and technical questions involving the behavioral and social sciences and, on occasion, takes the initiative in proposing studies on topics of national concern. Currently, the Assembly is responsible for supervising the activities of approximately 30 major committees and panels, involving over 500 distinguished behavioral and social scientists and other experts, and addressing a wide range of significant topics including national statistical programs, child development research and public policy, law enforcement and the administration of justice, population and demography, basic research in the behavioral and social sciences, human factors, vision and hearing, substance abuse, and aging. As you know, members of Academy committees serve without compensation, other than travel and other necessary expenses.

The Assembly's Committee on Ability Testing was formed in 1978, to undertake a broad examination of the uses and potential abuses of standardized ability tests in American society. Support for the Committee's work was provided by the Carnegie Corporation of New York, the National Institute of Education, the National Institute of Mental Health, the Office of Personnel Management, and the Ittleson Foundation.

Given its broad mandate, a special effort was made at the outset to appoint as members of the Committee a diverse group of outstanding scientists and other experts, a majority of whom had not been involved in existing controversies about tests. The Chairman of the Committee, Professor Wendell Garner, is a distinguished experimental psychologist, a member of the National Academy of Sciences, and former Dean of the Graduate School at Yale University. (His schedule, unfortunately, did not permit him to be present this morning.) Other members of the Committee include: Burke Marshall, former Solicitor General of the U.S. and now Professor of Law at Yale University; Philip Kurland, Professor of Constitutional Law at the University of Chicago; John Tukey, Professor of Statistics at Princeton University; Sociologist William Sewell of the University of Wisconsin and Chairman of the recent National Commission on Research; Marcus Alexis, former Commissioner of the Interstate Commerce Commission; and historian Oscar Handlin. Two distinguished psychometricians, Lee Cronbach of Stanford University and Melvin Novick, University of Iowa also served on the Committee. A complete list of Committee members will be submitted as part of Alexandra Wigdor's subsequent testimony.

The Committee has completed its report, which is now undergoing final editing prior to its formal publication by the National Academy

Press. In light of the report's significance and the complexity of the issues it addresses, the full report will be released to the public on February 2nd, 1982, at a day-long symposium sponsored by the National Academy of Sciences. I and the Academy's new President, Dr. Frank Press, hope that you and/or members of the Committee staff will be able to attend. (We will be in touch with your staff regarding specific plans for this important event.)

Alexandra Wigdor, Study Director for the Committee, is with me this morning to offer substantive testimony on behalf of the Committee. Sandy is a legal historian, trained at the University of Maryland, and she has directed staff work for the Committee during the last year and a half.

Finally, let me add a word of personal perspective on this project and on the legislation being considered by this subcommittee. For the first half dozen years of my career as a sociologist, from 1961 to 1968 I participated in the series of studies conducted by the Russell Sage Foundation on the social consequences of standardized testing. In addition to authoring two of the several volumes that resulted from this work, I was responsible for preparing the Foundation's report: Guidelines for the Collection, Maintenance, and Dissemination of Pupil Records, which subsequently served as a basis for what is now known as the Buckley Amendment. It was clear to us at the Foundation in the early 1960's that the problems associated with the widespread use of standardized tests in our society transcended the many technical issues of validity and test bias that were being argued then and that have been so hotly debated in the intervening years. At the heart of the controversy that has surrounded testing are the central questions of what qualities, talents, and skills the society values most and how best to nurture them. As John

Gardner put it in 1958, "the greatness of a nation may be manifested in many ways in its purposes, its courage, its moral responsibility, its cultural and scientific emminence, the tenor of its daily life. But ultimately the source of its greatness is in the individuals who constitute the living substance of the nation."* If this is true, then the means that are used to differentiate one individual from another and to reward those attributes most highly valued will determine not only the fortunes of individuals, but also that of the society. It is my hope and expectation that the Academy's forthcoming report will make a significant contribution to what will, and should, be a continuing national debate on these issues.

With your permission, Mr. Chairman, I should now like to introduce Alexandra Wigdor, Study Director of the Committee on Ability Testing. Following her statement, we are prepared to respond to questions that members of the Subcommittee may have.

---

*The Pursuit of Excellence: Education and the Future of America, Special Studies Project Report V, copyright by Rockefeller Brothers Fund, Doubleday and Company, Garden City, NY 1958, p. 1.

Mr. WEISS. Thank you very much. Miss Wigdor.

STATEMENT OF ALEXANDRA WIGDOR, STUDY DIRECTOR, COMMITTEE ON ABILITY TESTING, NATIONAL ACADEMY OF SCIENCES, WASHINGTON, D.C.

Ms. WIGDOR. Thank you. Thanks, Dave. Since Dave has already introduced you to the committee and its mandate I'll get right on with my remarks about the bill before you and about the Committee's work.

As I understand it, your primary interest is the use of standardized tests for admissions to colleges, universities, and professional schools. You'll obviously not be surprised to learn that the committee found that many junior and community colleges, most 4-year colleges and virtually all graduate and professional schools require applicants to submit standardized test scores to admissions officials. It is somewhat surprising to learn, however, that the great majority of undergraduate institutions are not very selective, that test scores and indeed GPA's are likely to present barriers only to those who rank very low among high school graduates and to that small portion of applicants who want to attend very highly selective and prestigious schools.

The situation in professional school admissions is very different. On suspects that the public perception of admissions test as being the great makers and breakers of educational opportunities result from the situation facing law schools and medical schools, where it's not uncommon to find 5,000 applicants competing for 120 places.

Despite the evidence that most undergraduate institutions are not very selective. The committee felt that many of the impulses expressed in the movement for test disclosure are well founded. Openness in the exercise of power is important to the American concept of equity. It seems a reasonable application of that principle to take the view that the allocation of educational opportunity is an exercise of power and that it ought not to be conducted entirely behind closed doors.

To place the issue of test disclosure in perspective, it is worth noting that it was not until 1958 that the College Board decided to report scores to students. Prior to that time scores had only been offered to the admissions officials, not to the students. That was the beginning step in what has become a bigger and bigger movement toward openness. The Buckley amendment was a much greater step in that direction. The committee felt that the truth it in testing movement is in the spirit of these developments. It is an important assertion of the interest of students and of society in general in the allocation of educational opportunity.

As a general principle, the committee believes that it is desirable that the entire post secondary selection process be open and known. In addition, there are very specific reasons for encouraging test developers to make more information available about their testing programs than has traditionally been the case. One of the biggest sources of problems with testing in America in the committee's judgment lies with the misunderstandings and the misinterpretations of test results by test users, test takers, and the public in general. The committee concludes in its report that well-developed and well-researched admissions tests can be useful selection aides if admissions officials understand their limitations. Like the sponsors of this bill, however, the committee realizes that the usefulness depends on having the knowledge, knowledge about the nature of the test and what it purports to measure, knowledge about the scaling methods used to give scores meaning and particularly knowledge about the limitations of tests.

Insofar as the bills before you then stem from a desire to promote more knowledge about tests and about the appropriate use of test scores, you'll find the committee's reports to be consistent with your efforts. The committee has been fully cognizant of the need to improve public understanding of this widespread technology and it expects its report to be an important source of information to test users, test takers, and indeed policymakers.

While the committee agrees with this need for information, however, it found the question of how openness can best be achieved extremely difficult. Three important questions present themselves. First, will disclosure result in substantial benefits to test takers? Second, will disclosure of test forms have dilatorious effects on the validity and reliability of tests? And third, is regulation by the Federal Government the best means to promote openness?

The first and most important question is whether full disclosure will bring substantial benefits to test takers, and, if so, to which test takers. The committee is skeptical about the assumptions and expectations expressed by many supporters of truth-in-testing bills. Two claims have been made: the first of them deals with test quality, the second with test fairness. Supporters have argued that such laws will provide an external spur to quality assurance by allowing test takers to judge the quality of the tests being used to make decisions about them. In the committee's opinion, this is highly unlikely prospect. Most test takers do not have the expertise among other things, the statistical expertise required to make any serious kind of analysis of the adequacy of the test or its research base. This is not to say the disclosure will not reveal ambiguities in test items, errors in scoring, we all know that there have been three or four instances of that already and to that extent it's been very useful. But in terms of a deeper analysis of the value of tests and the research that lies behind them, it's a highly unlikely prospect that disclosure will help. The more likely aspect is that wider access of the research community to the data basis will be useful and in that sense opening up the data basis to the research community might indeed aid the improvement of tests.

The committee did hear complaints by various parties that the testing companies have not always been as open as they might be to bona fide researchers and insofar as that is the case, the committee strongly recommends that the testing companies revise their policies.

Supports of truth-in-testing laws have also expressed the hope that disclosure will somehow make the testing process fairer, particularly that it will be fairer and to the benefit of disadvantaged minority students. The Educational Testing Service has released some data about the first two administrations of the SAT since the New York law that casts some doubt on that hope, it turns out that most of the students that asked for copies of the tests afterward and their scores or their answer sheet reported significantly higher family incomes, this is self-reported obviously, but nevertheless they were falling in the higher group and they had higher scores on the tests. Given this early and very partial data it doesn't seem likely that the hope that disclosure is going to help disadvantaged students is a very real and accurate hope.

Dave Goslin brought to my attention a study of "Sesame Street" that adds some corroboration to my fear—the committee's fear—that is more advantaged students will be the ones who benefit most. Researchers found that "Sesame Street" was indeed valuable to disadvantaged children who watched it but it was more valuable to middle and upper class middle children, they watched it more regularly, they had more reinforcing experiences in their daily lives, they got more out of it. One suspects that that might also be the case with disclosre.

While the committee was not convinced that full disclosure, therefore, would benefit the students and particular disadvantaged students, was equally unconvinced that the disclosure was going to bring about the demise of standardized testing. You've heard testimony from a lot of testing companies and I think over the last couple of years that testimony has changed a bit. Nevertheless, some of the companies remain convinced that disclosure will destroy the validity and the reliability of the their tests. Others have been much more optimistic about the likely effects and feel that they are going to be able to develop new aquating techniques, develop sufficiently large item pools and so on tha they will in other words, be able to rise to the challenge.

The committee felt that the experience of the last 2 years seems to bear out the more optomistic predictions. Largely because of the pressure brought by truth-in-testing supporters, we have a number of experiments going on right now. You're all familiar with the California and New York laws, there are partial and full disclosure experiments underway. The client boards of three and as I understand it now four of the major admissions test have decided to disclose their test nationwide. Educational Testing Service has instituted public interest principles by which they will commit themselves in their relationships with colleges, universities and their client boards to making more information available and otherwise being more open.

Apparently, the College Board is going to introduce new equating techniques with the SAT in 1982 that ought to solve some of the problems or at least experimentally they'll see whether they solve some of the problems the current techniques face with disclosure.

It's clear that so far none of these innovations has destroyed testing, what isn't yet clear and we'll only know in time is whether test standards can be maintained. Whether the realiability and validity of these instruments can indeed be maintained or indeed whether they'll be improved through disclosure.

With regard to the final question that I raised, there are those that claim that only Government regulation of testing companies can insure openness. Proponents of Governemnt oversight of the industry have drawn analogies to various regulatory precedents. Some use the analogy to public utilities, that the testing companies have a monopoly as public utilities do and, therefore, ought to be regulated. Some draw on the example of sunshine laws to justify opening up the process to public scrutiny. Some use truth-in-lending, truth-in-advertising models for the protection of consumer interests.

Committee on ability testing did not find these arguments totally convincing. First, the committee did not see evidence of systematic abuse by the testing companies serious enough to warrant Government intervention or at least to make it imperative.

Second, the analogies are an important respect misleading. There are three participants in the testing process, that is the testing company, the educational institution and the student. The bills before you reach only one of those participants, the testing companies. The actual decision-makers, that is the educational institutions, will not be reached by this legislation and therefore, the actual use of tests will not necessarily be effected in anyway.

Thus while supporting the general principle of openness, the committee was not convinced of the desirability of Federal legislation.

In sum, in the committee's judgment, none of the three basic questions raised by H.R. 1662, that is will disclosure benefit test takers, will disclosre compromise the tests and is Federal regulation the best route to openness, none of these questions now allow clear-cut answers.

This led the committte on ability testing to propose as its major recommendation of test disclosure a policy of watching and waiting. This recommendation is not a veiled proposal to do nothing, but it is a call to a different plan of action than you are not considering. Instead of speculating about the technical costs and consumer benefits of disclosure, the committee believes that the States, Congress, and the research community should monitor the developments of the next few years so that all parties can made a better informed judgment about creating a workable balance between openness and testing and objective assessment of academic ability.

This committee further recommends that the Department of Education through the National Institute of Education embark upon a positive program of monitoring the experience with test disclosure that is now underway. Because the conditions exist at present for an impirical investigation of the affects of various kinds of disclosure plans, NIE has the perfect opportunity and with your encouragement would no doubt seize the opportunity to underwrite the collection and evaluation of data needed to provide the basis of future policy.

That ends my formal remarks, we will be delighted to entertain questions.

[The prepared statement of Alexandra Wigdor follows:]

PREPARED STATEMENT OF ALEXANDRA K. WIGDOR, ON BEHALF OF THE COMMITTEE ON ABILITY TESTING, NATIONAL ACADEMY OF SCIENCES

The Committee on Ability Testing was convened in 1978 at a time of widespread controversy over the use of standardized tests in the schools, for college and university admissions, and for the selection of employees. Its central task was to conduct a broad examination of the role of testing in American life. Because the charge to the Committee called for a study of testing from a social perspective, the members have been sensitive to the need to go beyond questions of technical adequacy, and to explore the implications of test use for individuals, minority groups, institutions, and society as a whole.

The report of the Committee, which will be published in February by the National Academy Press, might best be characterized as a white paper on testing. It is intended to describe accurately the theory and practice of testing; to illuminate competing interests in a balanced fashion; and, ultimately, to help those who make decisions with tests or about testing to reach better-informed judgements than is now the case. The report is not a document written primarily by or for psychometricians and testing specialists; it represents the consensus of a multidisciplinary group of social scientists and legal experts, a majority of whom have no connection with testing, and it is addressed to policy makers and test users--lawmakers, judges, educators, personnel administrators, and the journalists and commentators who do so much to shape public understanding of testing issues.

My remarks today will focus on what I understand to be your primary interest, that is, the use of standardized tests in selecting applicants for admission to colleges, graduate schools, and professional schools. You will not be surprised to learn that the Committee found that many junior and community colleges, most four-year colleges, and virtually all graduate and professional schools require applicants to submit scores on ability tests for the use of admissions officials. It is surprising to learn, however, that the great majority of undergraduate institutions are not very selective, and that test scores (and GPAs) are likely to present barriers only to those who rank low among high school graduates and to that small group of applicants who want to attend the most selective colleges and universities--for the most part, the elite private schools. The situation in professional school admissions is, of course, very different; it is probable that the public perception of admissions tests as the great makers and breakers of educational opportunities derives much from the situation facing law schools and medical schools, where it is not uncommon to find 5,000 applicants competing for 120 places.

Despite the strong evidence that most undergraduate institutions are not very selective, and that test scores play a significant role only for a very small portion of students, the Committee felt that many of the impulses expressed in the movement for test disclosure are well founded. Openness in the exercise of power is important to the American concept of equity. It seems a reasonable application of the principle of openness to take the view that the allocation of educational opportunities should be recognized as an exercise of power that ought not be conducted entirely behind closed doors.

To place the issue of disclosure in perspective, it is worth noting that before 1958 test results were reported only to designated college and university admissions officers. In that year, the College Board decided to inform test takers of their SAT scores. The so-called Buckley Amendment of 1974, which requires educational institutions receiving federal financial assistance to allow students (or their parents or guardians) access to their academic files, represents a further step in the direction of openness with regard to information used to make educational decisions. The truth-in-testing movement is in the spirit of these developments. It is an important assertion of the interest of students, and of society in general, in the allocation of educational opportunity.

As a general principle, the Committee believes that it is desirable that the entire post secondary selection process be open and known. In addition, there are very specific reasons for encouraging test developers to make more information available about their testing programs than has traditionally been the case. One of the biggest sources of problems with testing in America, in the Committee's judgment, lies with the misunderstandings and misinterpretations of test results by test users, test takers, and the public in general. The Committee concludes in its report that well-developed and -researched admissions tests "can be useful selection aids if admissions officers respect their limits." But, like the sponsors of H.R. 1662, the Committee recognized that the usefulness of tests depends in large measure upon knowledge possessed by the participants in a selection decision (both the user of test results and the test taker)--knowledge about the nature of the test and what it purports to measure, the scaling methods used to give scores meaning, and the limitations of the test.

Insofar as the bills before you stem from a desire to promote more knowledge about tests and more appropriate use of test results, you will find the Committee's report to be consistent with your efforts. The Committee has been fully cognizant of the need to improve public understanding of this widespread technology. and expects its report to be an important source of information for test users, test takers, and policy makers.

While the Committee agrees with the sponsors of the bills under discussion about the need for more generally available information about the characteristics, appropriate uses, and limitations of standardized tests, however, it found the question of how openness can best be achieved very difficult. Three basic questions present themselves:

1. Will disclosure result in substantial benefits to test takers?
2. Will disclosure of test forms shortly after some or all administrations have damaging effects on the validity or reliability of the tests?
3. Is regulation by the federal government the best means to promote openness?

1. The first and most important question is whether full disclosure, that is, disclosure of the actual test questions and answers, will bring substantial benefits to test takers, and if so, to which test takers. The Committee is skeptical about the assumptions and expectations expressed by many supporters of truth-in-testing legislation. Two claims have been made: one concerns test

quality and the other test fairness. Supporters have argued that such laws will provide an external spur to quality assurance by allowing test takers to judge the quality of tests used to make decisions about their educational future. However, expertise far beyond the capacity of most test takers is required for any serious analysis of the adequacy of a test and its research base (although disclosure may be, and occasionally has been, useful in turning up ambiguities in specific items and errors in scoring). Wider access of the research community to test data is far more germane to improving the quality of tests. There have been complaints that the testing companies do not always welcome bona fide researchers from outside the organization. Insofar as this is the case, the policies ought to be revised.

Supporters of truth-in-testing laws have also expressed the hope that disclosure will somehow make the testing process fairer, and particularly that it will benefit disadvantaged and minority students. On the question of which students will benefit from test disclosure, an Educational Testing Service staff analysis of the first two SAT administrations since the New York law took effect indicates that the students who requested a copy of the test booklet and answer sheet had significantly higher mean scores and higher (self-reported) family incomes than those who did not. The hope that disclosure will improve the competitive position of disadvantaged and minority students is not borne out by this early evidence, although it is just that--very early and partial evidence. Some corroboration is offered by the study of the children's television program, Sesame Street, published by the Russell Sage Foundation in 1975. The creators of the program intended that it enrich the early learning experiences of disadvantaged children, which it did. But the infant viewers from middle- and upper middle-class homes benefited even more because they watched the program more regularly and had more reinforcing experiences in their daily lives. The experience with Sesame Street suggests that while full disclosure may be of some benefit in terms of student performance, it is likely that the best students will receive the most benefit.

2. While the Committee is not convinced that full disclosure would be of much benefit to the the student, either in terms of test quality or in improving the performance of disadvantaged test takers, it was equally unconvinced that full disclosure would cause serious problems with regard to test validity and reliability in large testing programs. As you are well aware, many test developers have expressed concern that test disclosure will have negative effects on the quality of the tests. Some feel that the pool of possible items is so limited

and item development so difficult a process that disclosure after each administration of a test will soon give test takers what amounts to prior knowledge of the test questions. They have also expressed the concern that the reliability of the tests, which relates to the consistency of test scores, will be reduced because full disclosure rules out the traditional equating techniques. But there are other test developers who feel that the industry will be able to respond to these challenges, that they will be able to develop new item pools and new equating techniques rapidly enough to maintain test quality.

The experience of the last two years seems to bear out the more optimistic predictions. Largely because of the pressure brought by supporters of open testing, experiments with various kinds of test disclosure have been set in motion. California and New York have passed state "truth-in-testing" laws. The California law requires that facsimiles of postsecondary admissions tests, along with validity data and scoring information, be filed with the Postsecondary Education Commission. Test takers are to be given information about the purposes of the tests, the nature of the subject matter, scoring procedures, and sample questions, thus placing a legal obligation on test publishers where formerly voluntary action sufficed. The New York law requires full disclosure. Under the terms of the Lavalle Act, test producers must file with the Commissioner of Education and supply to the test taker upon request the actual test questions and answers within 30 days of administration.

In addition to these state disclosure laws, some of the testing companies and client boards, despite continued concern about the effects of full disclosure on test quality, have begun to respond positively to the challenge. The Educational Testing Service has developed "public interest principles" to guide its staff and its relations with the client boards and colleges and universities that use its admissions tests. Of more immediate consequence, three client boards, the Law School Admissions Council, the Graduate Management Admissions Council, and the Graduate Record Examination Board, have decided to disclose their examinations nationwide. And new equating techniques are being developed and apparently will be used on an experimental basis with the SAT in 1982.

None of these innovations has brought about the immediate and total collapse of standardized admissions testing, although the long-term effects for good or for ill remain obscure.

3. With regard to the final question, there are those who claim that only government regulation of the testing companies can ensure openness. Proponents of governmental oversight of the industry have drawn

analogies to various regulatory precedents: some argue that testing companies have a monopoly on the instruments of human resource allocation that would justify their treatment as a public utility; some draw on the example of sunshine laws to justify opening up the process of test development and validation to public scrutiny; others believe that "truth in lending" and "truth in advertising" provide models for the protection of consumer interests.

The Committee did not find these aguments totally convincing. First, the Committee did not see evidence of systematic abuse by the testing companies serious enough to make government intervention imperative. Second, the analogies are misleading. There are three participants in the testing process: testing company, educational institution, and aspiring student. The proposed test disclosure bill does not directly influence the behavior of the actual decision maker, the educational institution, and, therefore, will not necessarily have much effect on how tests are used. Thus while supporting the general principle of openness, the Committee was not convinced of the desirability of federal legislation. Nor did it find self-evident the efficacy of the laws currently in operation or under discussion in the states for significantly improving tests or the way they are used.

In sum, in the Committee's judgment, none of the three basic questions raised by H.R. 1662--Will disclosure benefit the test takers?; Will disclosure compromise the tests?; Is federal regulation the best route to openness?--allow clear-cut answers at present. This led the Committee to propose, as its major recommendation on test disclosure, a policy of watching and waiting. This recommendation is not a veiled proposal to do nothing. But it is a call to a different plan of action than you are now considering. Instead of speculating about the technical costs and consumer benefits of disclosure, the Committee believes that the states, Congress, and the research community should monitor the developments of the next few years so that all parties can make a better informed judgment about creating a workable balance between openness in testing and objective asessment of academic ability.

The decision of the three client boards to disclose their tests nationwide, together with the existing state open-testing programs in New York and California, provides all parties to the truth-in-testing debate with the opportunty to establish the effects of disclosure empirically. The experience with these efforts over the next few years will indicate whether full disclosure is technically feasible, what the financial costs will be, whether there are improvements in the quality of admissions tests, and whether the hoped for benefits accrue to sudents.

The Committee further recommends that the Department of Education, through the National Institute of Education, embark upon a positive program of monitoring the experience with test disclosure. Because the conditions now exist for an empirical investigation of the effects of various plans for test disclosure, NIE has the opportunity--and with your encouragement would no doubt seize it--to underwrite the collection and evaluation of data needed to provide the basis for future policy.

Mr. Weiss. Thank you very much. I think there had been a request for Dr. Goslin that the list of the members on the committee on Ability Testing be entered into the record.

Ms. Wigdor. Yes.

Mr. Weiss. That will be done at this point.

[Information submitted by Alexandra Wigdor follows:]

COMMITTEE ON ABILITY TESTING

WENDELL R. GARNER (*Chair*), Department of Psychology, Yale University

MARCUS ALEXIS, Department of Economics, Northwestern University

WILLIAM BEVAN, Department of Psychology, Duke University

LEE J. CRONBACH, School of Education, Stanford University (psychometrics)

ZVI GRILICHES, Department of Economics, Harvard University

OSCAR HANDLIN, Director of Libraries, Harvard University (history)

DELMOS JONES, Department of Anthropology, City University of New York

LYLE V. JONES, L.L. Thurstone Psychometric Laboratory, University of North Carolina (psychometrics, statistics)

PHILIP B. KURLAND, Law School, University of Chicago

BURKE MARSHALL, Law School, Yale University

MELVIN R. NOVICK, Lindquist Center for Measurement, University of Iowa (statistics, psychometrics)

LAUREN B. RESNICK, Learning Research and Development Center, University of Pittsburgh (psychology, educational psychology)

ALICE ROSSI, Department of Sociology, University of Massachusetts

WILLIAM H. SEWELL, Department of Sociology, University of Wisconsin

JANET T. SPENCE, Department of Psychology, University of Texas

ALAN A. STONE, Law School, Harvard University (psychiatry, law)[*]

MARY L. TENOPYR, Human Resources Laboratory, American Telephone and Telegraph Company, Morristown, N.J. (industrial psychology)

JOHN W. TUKEY, Bell Telephone Laboratories, Inc., Murray Hill, N.J. and Princeton University (statistics)

E. BELVIN WILLIAMS, Educational Testing Service, Princeton, N.J. (psychology)

ALEXANDRA K. WIGDOR, *Study Director*

SUSAN W. SHERMAN, *Senior Research Associate*

GLADYS R. BOSTICK, *Administrative Secretary*

---

[*] *Member until 1980.*

Mr. WEISS. Could you describe for us the methodology that was used by the committee in coming to its conclusions?

Ms. WIGDOR. Yes, indeed. I don't think I'd call it a methodology, but——

Mr. WEISS. The process.

Ms. WIGDOR. The process, yes. The committee's investigation involved a number of kinds of strategies and in the first place we did hold our own hearings 2 years ago, at which time we had representatives of testing companies, test users, student associations, various advocacy groups, some of which for example, represented minority interests, the Urban League was there, NAACP, and the Association of Black Psychologists. I can't remember the full list at the moment. But, at these hearings we asked people both to give us their impressions about standardized testing in general and about specific issues. Some of the people did speak on test disclosure

In addition to that, we invited written submissions from—as I remember 175 interested individuals and organizations and compiled a huge library of their experiences with testing, their problems with testing, so on and so forth. That was one kind of activity that was aimed at trying to allow everybody who had an interest in testing issues in general and a disclosure and specific to make sure that the committee was aware of their position.

Another kind of activity included marshaling the research literature. Now, Academy committees do not themselves do original research, what they do is bring together and summarize the best existing research. This is done through literature searches through various kinds of connections with scholars working in the field, so on and so forth.

As part of its information-gathering effort, the Committee also leaned upon the services of experts outside its own membership. We have as part of the report—appended to the report a collection of 11 signed papers, a couple of them done by Committee and staff people, most of them done by outside experts, one of the oncollege admissions.

These papers again brought together the best of the research for the Committee's consideration. Committee membership itself, of course, included years and years and years of experience and expertise on various kinds of issues, not always very obvious.

For example, I think both of the members of the committee who are on law school faculties have sat on law school admissions boards. Many of the members have had that sort of dreadful experience so that they had personal experience with the problems facing admissions officers and personal experience of how the process really works, how the decisions really are made, what role scores play. In other words, that kind of firsthand knowledge.

Dave, have I missed any of the obvious ways in which our committee got it——

Mr. GOSLIN. The obvious way that you've missed is that the Committee met many times.

Ms. WIGDOR. Yes.

Mr. GOSLIN. I'm sorry the most obvious thing that Sandy left out is that the Committee met many times over a 3-year period and—

Mr. WEISS. By many, how many times would you say?

Ms. Wigdor. I think we had a total of 12 meetings, one of them being a 4-day meeting, most of the rest being 2½-day meetings. And that was of the full committee. In addition to that, the committee in the last 1½ year broke up into working groups, writing groups, four or five committee members got together in three different writing groups and they met on an average of three times, the writing groups and then back into the full Committee, so it was a great deal of contact. Sharing of knowledge of information, working out positions, working out the wisdom that——

Mr. Weiss. And were there position papers on specific issues presented to that group with——

Ms. Wigdor. Yes.

Mr. Weiss. Proposed options for judgment or conclusions to be adopted and then there was a vote taken? Did the committee ultimately come to a decision that they would say yes or no as to a particular conclusion?

Ms. Wigdor. Yes; we used the technique of position papers to start getting the committee to form basic opinions and then as the writing groups got to work, drafts were circulated among the whole committee, everybody responded in writing to the various drafts, things were redrafted, where there was difference of opinion more information was gathered, more discussion was held until gradually we had what is a genuine document of consensus.

Now, this is not to say that every member of that committee agrees with every single word in the report; you never get 18 people to agree to every word of anything.

Nevertheless, it is a document that represents the consensus opinion and I think a genuine consensus. We have had an unbelievable amount of participation of the committee members in the writing process and in working out positions and policies.

Mr. Weiss. And the whole process took how long?

Ms. Wigdor. The first meeting was held February 3 and 4, 1978; it's been a long process.

Mr. Weiss. And you will have your report released February of 1982, right?

Ms. Wigdor. That's right.

Mr. Weiss. Roughly 4 years from beginning to end of the process.

Ms. Wigdor. That's right. I might also add that—this—is true of any Academy report, but it's true of ours that it went through a double level of review as all Academy reports do. It's reviewed by the Assembly of Behavioral and Social Sciences, of which Dave is the executive director, that review involved six members of the assembly who read the report and wrote written—gave us written comments to which we responded. Thereafter, the report went through review by the Report Review Committee of the National Academy of Sciences. These reviewers were not primarily social scientists, but scientists who as a general rule look at the adequacy of evidence so on and so forth and we had very helpful comments at that level of review as well.

That process assures that sloppiness doesn't slip through, that the committee does indeed have the documentation to support its position. So on and so forth. We found it a very useful process.

Its had a good deal, in other words, a good deal of peer review.

Mr. **Weiss**. So that's it, again, just to be sure that I am correct. This was not any kind of independent research, original research, that the Academy undertook. It was really a study based on information gathered by it and submitted to it, with the original work having been done by others and out of all that——

Mr. **Goslin**. Including the members?

Ms. **Wigdor**. Including, yes.

Mr. **Weiss**. Including the members and out of all of that, a consensus document was arrived at which takes in rough form the position as outlined in the testimony that you gave us.

Ms. **Wigdor**. On disclosure, which is just one of a very large number of issues, yes, but this is the position of the committee on disclosures.

Mr. **Weiss**. On disclosure, right.

Ms. **Wigdor**. For college admissions or postsecondary admissions. I might——

Mr. **Weiss**. Did you find that the committee's positions changed as a result of developments that occurred during the interim of the study? As of 1978 there was no legislation on the books anyplace. Legislation here had not been introduced and the conclusion that you've come to now that we've got legislation on the books is "let's watch and see how it works out and let's monitor it closely."

Ms. **Wigdor**. Indeed that, yes, indeed that is the case.

One of the reasons that this project took 3½ years is that so many activities suddenly were underway and the committee felt that since its mandate was to look at testing from a social perspective, social impact of testing, that it must indeed be cognizant of what's going on in society. We have been at your hearings for the last 2 years. We have been following these developments very closely. Obviously when the committee was first formed they did not know they would be talking about test disclosure. Yes, they have been monitoring these developments.

Mr. **Weiss**. The committee did not know that it would be talking about test disclosure?

Ms. **Wigdor**. Did not know that test disclosure was going to be an important part of its report.

Mr. **Weiss**. OK. That sort of obviates my next question which was going to be, was there then a change in, in position from February 1978 to sometime in 1980, sometime after the time when the New York legislation was adopted, for example? Had there been a change of position on the part of the members of the committee?

Ms. **Wigdor**. That's really a question I can't answer, at least I think in the way you want it answered. Yes, in one sense there is always change when you've got 18 people with different kinds of experiences and expertise getting together and educating one another. Yes, there is change. I would say that half the members of the committee, those who are not testers, psychologists or in some other way connected with or having experience of testing, they didn't know disclosure was an issue to begin with. They knew it by the time things were moving here and in New York. In that sense, of course, there was change. People were educated. I would say yes, opinion evolved. I think that the questions, the fundamental questions, have remained pretty much the same. That is to say they were not sure in the beginning what the effects of full disclosure would be on tests and they are still not sure of that. They have not found, there has not been enough time for enough evidence to be in to make that kind of assessment.

So in that sense, that question remains and I would say the other question, that is, is this really going to benefit students and, particularly, is it going to benefit disadvantaged students? That question still remains as the central question and the committee feels it has no real answer yet.

These things have remained constant overtime and the committee found them very troubling questions.

Mr. GOSLIN. At the same time, I can remember the first meeting of the committee and there was a very clear commitment on the part of members of the committee to the concept of openness.

Ms. WIGDOR. Yes.

Mr. GOSLIN. And that has remained constant and is reflected clearly in the committee's report.

Mr. WEISS. Well, I wonder if you would expand on that a little bit. I find a seeming conflict or perhaps contradiction in the commitment to openness, except in the implementation of steps which would in fact lead toward greater openness.

Ms. WIGDOR. Obviously we'll have to define degrees and, well, define terms first, and talk about degrees of openness.

Mr. WEISS. OK, what do you mean by openness?

Ms. WIGDOR. By openness, not what I mean, but what I think the committee means by openness.

Mr. WEISS. Right.

Ms. WIGDOR. Would be something, I think you've even used the word yourself, getting away with—doing away with the mystery.

Mr. WEISS. Doing with——

Ms. WIGDOR. The mystery that has——

Mr. WEISS. Yeah, right.

Ms. WIGDOR. Sorry. That has traditionally surrounded tests. There was, I think, a carryover from the early days of testing when the expert was put on a pedestal and so on. There has been a tendency in our society to shroud these instruments in mystery.

The committee, I think, feels that that is extremely detrimental. This is a too widely a used technology to be allowed to be thought of and used that way. So the committee is entirely in agreement with you that we've got to do away with the mystery that test users and test takers simply need to know a lot more about tests than they do.

I think we are, the committee is not convinced, in terms of your position, is that the ultimate step, that is making the answer, the questions and answers available afterward, not convinced that this is an absolutely essential step to take. Not convinced that it's not.

Mr. GOSLIN. It's not convinced that it isn't either.

Ms. WIGDOR. Right. But, just, you know, that is where they then draw the line, of—there is still an unaswered question here in terms of balancing openness against the technical adequacy of the tests.

If full disclosure, if after how ever many years it takes to see what kinds of new equating techniques the companies can come up with, and so on and so forth, if it turns out that disclosure is not having any detrimental effect on the test, the committee would say unanimously, of course, make them available. If nothing else it will give the students a little peace of mind.

If, on the other hand, it would destroy the usefulness of the test, the committee would say you should draw the line before that. Make facsimiles available, make much more information about scaling and what scores mean and the limitations of tests. Make much more of that kind of information available, but draw the line on full disclosure if it will ultimately destroy the quality of the tests.

Mr. **Weiss**. Let me cite two examples for you. You're familiar with them. The law college admissions council has decided to go the full, total route of openness. The medical college admissions group has gone 180 degrees the other way. They've gone to court to prevent the New York law from applying to them. They had Dr. Cooper testifying yesterday that the information that they test is so limited and so finite that they have limited number of questions that they can ask and, therefore, disclosing their tests would in fact do harm to the entire testing process.

Now, these are two professional organizations, that is organizations dealing with two of the major test-taking factions in this country, coming to those conclusions.

I assume that you looked at the approaches that are taken by those two professions and I wonder if you have drawn any conclusions at all from the varying positions which they take.

Ms. **Wigdor**. The committee did not and does not in its report support one or the other of those positions. It does, I don't think it does in the report, but it did in its discussions, point out a basic difference between the situation of the Law School Admissions Council and their test and the medical exam and that is the law school test is not as curriculum bound. There is no common pre-law undergraduate education as there is premed. Therefore, their test is designed to be less curriculum bound than the medical school test is. Where that test, as I understand it, I'm not speaking as an expert, their test is much more dependent on having had chemistry and having had zoology and having had biology and so forth. It's much more bound to a specific premed curriculum.

That means that their concerns could be justified and the other position could be justified as well. I'm not an expert and I'm not really in a position to say. The committee did not excel.

I would say only that there might be something there that both, both positions could be justified. Not that either of them is because of the differences in the examinations.

First, one of the real problems in talking about testing is we talk about tests but there are tests and tests and tests and each test has its on character and its own uses and we're generalizing when sometimes the generalizing is not furthering our understanding.

Mr. **Weiss**. Without wanting to be at all pejorative——

Ms. **Wigdor**. Yes?

Mr. **Weiss**. It's sort of—in talking about these two groups, it puts me in mind of talking to two people, one of whom says the world is round and the other one of whom says the world is square and you come along and say, I guess either one of you could be right.

Ms. **Wigdor**. Depending on what world you live in.

Mr. **Weiss**. Well, I, I——

Ms. **Wigdor**. It's not that they're——

Mr. **Weiss**. I think, I think that that's why some of us, perhaps all of us, are going to have a problem with the report that you come out with.

Mr. **Goslin**. Another answer to that same question is that not enough time has passed yet to see what the potential results of the policies taken by those two organizations are. I think that there is a natural experiment going on. The law school people did one thing. The medical school people did another. And it will be used consistently with the committee's report that says let's wait and see what happens.

Mr. **Weiss**. Is there, any reason—I'll give you a ball right over the plate—is there any reason why, given the description that you gave us as to how the committee arrived at its conclusion, given the process that you have observed these subcommittees going through over the course of the last 2½ years, is there any reason to believe that the conclusion that you have come to is any sounder than the conclusion that we may come to?

Ms. **Wigdor**. Yes. If I may be so bold.

Mr. **Weiss**. Go ahead.

Ms. **Wigdor**. You have a great many things on your plate all the time. You cannot devote it, although your staff people, I'm sure, devote a great deal of their time to the problems of testing. You people cannot devote the kind of time that the committee was able to do to these questions, I think. I'm just assuming that you have so many responsibilities in this case.

Mr. **Weiss**. This committee has, in 2 years, met a dozen times. You in 4 years and your committee have met a dozen times. I assume that your people, that is the members of your committee, do other things than just devote themselves to the committee.

Ms. **Wigdor**. Yes; indeed we do.

Mr. **Weiss**. They have professional lives.

Ms. **Wigdor**. Yes.

Mr. **Weiss**. This is sort of a sideline for them.

Ms **Wigdor**. Yes; indeed they do.

Mr. **Weiss**. We are, for better or worse, almost professional at being Members of Congress and this, most of us, is the major committee assignment that we have.

Ms. **Wigdor**. Yes.

Mr. **Weiss**. Again, within that context that I asked as to, since you're not claiming any independent research or statistical objectivity or whatever, why your conclusion is any sounder than ours.

Ms. **Wigdor**. All right.

Mr. **Weiss**. And I don't know what ours will be. I know what mine is, but I don't know what the committee's will be.

Ms. **Wigdor**. Let me approach the answer differently. Obviously, my first thought was not a very good one. My first reaction to your question was obviously not a very good one. I think, perhaps, we can get a better answer by looking at the functions you are trying to make in policy.

The committee is trying, not to be Congress and make policy, but to evaluate on the basis of scientific evidence, questions about testing and then suggest, on the basis of their knowledge of the law, their knowledge of government, their knowledge of the society, the sociologists, possible policy alternatives for people like you, policymakers, to think about.

Those that came to the committee were experts. Each of them representing a different area of expertise and it was on that level that they talked. It was on that level that they gathered and looked at information and then came to the conclusions. I think that probably is the fundamental difference between their group process over the last 4 years and your group process over the same period of time.

Mr. WEISS. Well, it's also possible to suggest, and again I do not do this suggesting ulterior motive, but independence, it's also possible to suggest that since an awful lot of the people on the committee have had or continue to have some involvement with the very field that is being questioned and explored, whereas our people are not, to that extent, involved, that we perhaps may be more objective about it.

Mr. GOSLIN. This is a classic problem that faces every single committee of the National Academy of Sciences. No matter what controversial issue we are addressing and there are several issues.

One is how you separate the scientific evidence, the knowledge base, from the policy consideration, from the political considerations, and it is not in general the task of the Academy to do the job of the Congress. We are supposed to say, what do we know about a problem and we leave the political judgments to those persons elected to exercise that responsibility.

It is also always a problem in composing it's committees to find the experts that are both independent and expert and you can't get to be expert unless you are involved in a field and, therefore, you are no longer independent. And we have this problem with all Academy committees whether we are looking at the effect of fluorocarbons on the ozone layer. Dow Chemical Co. scientists know more about fluorocarbons then anybody else, because they make them, so we may want somebody like that on a committee and we certainly wouldn't have wanted to compose a committee on ability testing without having as a member of that committee some people who know a hell of a lot about testing.

We also feel that it is very important to try to balance those kinds of expertise with many other kinds and, indeed with some individuals who come with a very broad perspective, like Bert Marshall for example, and so that's always our dilemma and we try to do the best we can in putting together a committee that has a balance of perspectives and I can only leave it to you to judge from the list of members of the committee and from the report itself when it appears as to whether we have done a reasonable job.

Mr. WEISS. The suggestion that Doctor Wigdor made as to the scientific nature of the testimony that was presented—I think that's what you said—You know, we received much of the same testimony, I suspect.

Ms. WIGDOR. No, I did not.

Mr. WEISS. Well, how did scientific get into the last answer that you gave us?

Ms. WIGDOR. Sorry, I meant that there were members of the committee who were qualified. Statisticians, for example, who were qualified to do analyses of tests and present to the rest of the committees who did not have the quantitative abilities really expert opinions about the tests. That's what I mean by scientific. There were legal scholars who were in a position to present to the other committees who did not have their expertise expert information about the requirements of law or whatever. In other words, each of these people on the committee had some kind of scientific or other expertise with which he could bring or she could bring to the problems we are looking at. In that sense I meant scientific expertise.

Mr. WEISS. You know from observing these hearings that we too had some scientific testimony. On both sides of the issue, I might say, and many times those people were in diametric oppostion with one another.

Ms. WIGDOR. I beg your pardon?

Mr. WEISS. Pardon?

Ms. WIGDOR. I didn't hear your last question.

Mr. WEISS. I said some of them were in diametric opposition to one another and, again, I don't know the composition of your board. I don't know who the members of that committee are, but I would be surprised if it in fact included people who were very strongly questioning of the field of psychometrics and the whole testing area and, if, in fact they would come up with the kind of conclusion that you suggested. That is, of just watching and let it go by to see how it develops, because the impression that we got in the course of our hearings was that there really is a very serious amount of soul searching going on in the field at this point, on the part of people who have spent the bulk of their lives in the field, as to whether what they are doing is at all valid, whether it's constructive, all kinds of questions.

Ms. WIGDOR. I can assure you that within the membership of the committee, the opinion ranged very widely as to the value of tests. There were a number of committee members who considered them extremely valuable for selection, for assessment. There were other committee members who thought they were relatively trivial in the grand scheme of things.

Mr. WEISS. Somebody who said they were postively harmful?

Ms. WIGDOR. Yes. I'm trying to remember and be absolutely accurate. There were committee members who argued that position very strongly. I think what the committee did not have in it's membership were people who were essentially advocates and I think that the Academy tries for the very reasons Dave has discussing before, to avoid that kind of committee composition because we are aiming for reports that are dispassionate, that can weigh and balance and look at the whole situation.

In regard to the second part of your question, yes, there obviously is diametrically opposed opinion offered in part by people who are considered experts, in part by laymen, and I guess the committee's position, instead of just sitting in the middle, is saying, look, a lot of these questions simply cannot be answered yet and until you can answer them, solutions aren't known.

Mr. GOSLIN. Moreover, I don't see any reason why our committee and your committee can't come out in a different place. You're operating on the rationale for your decision may include some things that were not part properly of the rationale for our committee's proposal.

Mr. Weiss. All I was asking for was whether in fact there was any reason to consider your conclusions to be at a higher level than ours.

Mr. Goslin. At a different level. I wouldn't put higher or lower on that.

Ms. Wigdor. I would say there's certainly reason for our report to receive, at your hands, a good deal of serious consideration. I would not say that this, our report should be the only or one of the two or three determining factors in your decisions.

Mr. Weiss. I guess I have one substantive question. I thought that process was important to go into, especially since we don't have the report itself in front of us.

Ms. Wigdor. Yes.

Mr. Weiss. The title of your committee is, Committee on Ability Testing.

Ms. Wigdor. Yes.

Mr. Weiss. And we have had some exploration from time to time, not as to the word ability, but as to the word aptitude because that's what the SAT, the A in the SAT is aptitude, and I think in the course of these hearings, we've moved at least some extent in redefining what is meant by the word "aptitude," and I wonder how you defined the word "ability."

Ms. Wigdor. That name, that word was chosen, thank you, that word was chosen with some malice aforethought because the traditional words, "aptitude" and "achievement," are loaded with all sorts of baggage that destroys rational discussion at this point. "Ability" was chosen as a more neutral word and, in a sense, an umbrella word. The words "aptitude" and "achievement" were thought to be, originally I guess, two different kinds of things. One, native ability that was there in the genes and the other from what you learned in school.

The committee's position is that this is an unnatural distinction therefore, it uses the word "ability" it uses "achievement" and "aptitude," it couldn't avoid it entirely. It tries to make the point that these are words that talk about two ends of the continuum. No knowledge springs full blown from someone's brain. No knowledge is only a product of experience. It is a product of environment and the brain working together and that's what we mean by ability. Whatever the developed ability is at the moment of testing is what we are talking about. It comes from a multitude of sources.

Mr. Weiss. Thank you.

Mr. Erdahl.

Mr. Erdahl. Thank you, Mr. Chairman. Let me, Doctor Wigdor, just read from a couple observations you make in your statement. From these I will have a question and another observation.

On page 2 it said:

> It is surprising to learn,— however, that the great majority of undergraduate institutions are not very selective and that test scores GPA's are likely to present barriers only to those who rank low among high school graduates and to that small group of applicants who want to attend the most selective colleges and universities, for the most part, the elite private schools.

And then the other observation was toward the end of your statement on page 10:

The proposed test disclosure bill does not directly influence the behavior of the actual decision maker, the educational institution, and, therefore, will not necessarily have much effect on how tests are used.

I understand from Chairman Weiss' perspective, his bill does not attempt to get at any one or group of testing companies, but rather what impact the tests have on students. Obviously, the bill would have some impact on testing companies. I'll ask both of you, what in your educated judgment is the ultimate impact on the student or potential students to a university or institution of learning?

Ms. **WIGDOR**. Of full disclosure?

Mr. **ERDAHL**. Yes, right, of the bill. That's what I mean, the impact of the bill.

Mr. **GOSLIN**. You're getting back to the bill?

Mr. **ERDAHL**. Yes.

Ms. **WIGDOR**. I think, you know, you put the two statements together for a majority of applicants in undergraduate institutions; disclosure probably will not affect their educational future. They were going to be accepted anyhow.

Mr. **GOSLIN**. Or not accepted.

Ms. **WIGDOR**. Or not accepted anyhow, right. No, for the majority, they're going to be accepted.

For the minorities at the two ends—that is, the students who aspire to attend a highly selective school—I mean disclosure comes after the fact. I'm not sure, I'm not sure what the effect would be in either, for either of the groups for whom testing might make a difference.

If the assumptions of some supporters, perhaps Mr. Weiss, I don't know, if some supporters of disclosure are correct, if the assumptions are correct that test quality will improve because of disclosure, everybody benefits, and particularly those for whom the test score is a real criterion of admission.

If test quality declines, I guess everybody loses a little bit.

Mr. **GOSLIN**. The more a test taker knows about the requirements of the test and what kinds of skills he'll be required to demonstrate on the test, the better. The more likelihood that mistakes or errors can be corrected or there is an appeal, the better.

It seems unlikely to me that very large numbers of test takers, very large proportions, are going to ask for their, going to ask for their answers and full copies of the tests and so on. We haven't seen evidence that that——

Ms. **WIGDOR**. Well, there is——

Mr. **GOSLIN** [continuing]. Is widespread.

Ms. **WIGDOR**. No; there is, however, some interesting evidence on that. Apparently the number of students who will request their test forms is very much a function of how much they have to go through to request them.

Mr. **GOSLIN**. Like everything else.

Ms. **WIGDOR**. Right; I mean if you've got to pay your $4.25 and write a letter, not many are going to request. If on the other hand you can check off, when you walk into the place, more are going to request. So that's, that all remains to be seen. That's one of the questions that remains to be seen. Our people actually are going to ask for this information. If they do, is this somehow going to reduce test anxiety for the next generation or something. It's not obvious that it will have any dramatic effect, but it might.

Mr. **Erdahl**. OK. Another question and obviously you don't have to answer, but for reasons that the other witness is not here, you might not wish to respond to this, but I think both of you were here when the previous witness talked about the FTC's work on testing. Would either of you care to comment on that?

Ms. **Wigdor**. Only briefly and not as myself being an expert. I can only repeat some of the general conclusions that the committee drew about the coaching question.

One, they were not convinced, they were not satisfied with the kind of evidence that is available on coaching. The research studies are, have a lot of problems. The FTC data analysis had some problems and the others do, too. It's very hard to get comparable groups, some of whom have taken a test and some of whom have not, and compare scores. From a scientist's point of view, the data isn't very clean. Therefore, it's sort of hard to make any accurate assessment of the value of coaching.

On the other hand, when you start talking about long numbers of hours, you're talking about learning. The committee realized that there is a problem of semantics and ultimately came to the conclusion that although information is partial, the research base is not terribly good. It would be, since these tests are so widespread, it would be useful for schools, high schools, to make a modest effort to inform students, coach students if you will, on test taking techniques.

I mean clearly all students should know whether or not on a test facing them it pays to guess. Some students don't know that. So the committee felt that it would be important for high schools to make sure that the students go into the testing situation knowing the basics of test behavior.

In terms of any longer term coaching in the high schools, the committee felt that in so far as the content of any coaching instruction was real content just things they should be learning anyhow, reading and writing and arithmetic, fine, you're not going to lose anything. On the other hand, if it makes a real distortion of the high school curriculum, it's probably not a very good idea at all.

Mr. **Erdahl**. Interesting observation. It coincides with something we heard earlier today and this is not an official definition, but I guess mine. If you study for 4 or 5 hours before the test with some special people it might be coaching. If you do it for 30, 40 hours, it might be more of a learning experience. I agree that we get into a question of semantics, but you made the point about some basic things. I hope that our schools would stress them, whether to move along, don't skip, if you don't know—guess, these things that I think most of us had to do to survive tests.

Ms. **Wigdor**. Sure, and I think that the testing companies can also contribute a lot by making that very information much more explicitly available to test takers in their information packet. There is no reason not to instruct students in test taking techniques then. They may indeed it right now. I don't know. They didn't used to.

Mr. **Erdahl**. Thank you very much and thank you, Mr. Chairman.

Mr. **Weiss**. Thank you.

On the question of how important it is since the majority of students get into their college of the first choice, in any event, and it's only those who apply to the elite schools who might or might not be disadvantaged by not being, not taking the coaching courses or not having their test scores improved by the minimal number of points, that would in fact—don't you think that that's important as to whether in fact someone gets into one of the elite schools or not if that's what he or she has his or her heart set on.

Ms. **Wigdor**. It's important to that individual. The committee was not convinced that it's socially important enough to justify or require legislation and I think the committee was particularly bothered by two things that are present in the movement for truth in testing or test disclosure.

One, there seems to be a great hope that somehow the disadvantaged, the minority student, is really going to benefit from this and then this other thing, the privileged student, I mean the students we're talking about who want to go to the elite schools are by and large the privileged members of our society. The two impulses do not coincide if you see what I mean and I think the subcommittee would say—notice there is a hesitation—I think the subcommittee would say that the fate of the very privileged student is fairly well taken care of in this society anyhow and probably should not, that concern probably should not be the basis for national legislation.

Mr. **Weiss**. I'm very rarely identified with being overly concerned about the rights of the overly privileged, but I must tell you that I would be concerned if in fact we adopted policies or didn't adopt policies because they only affected the rights of the advantaged, and I'm not sure that you've really thought your position through as carefully as you would like to.

Ms. **Wigdor**. That's probably quite true. I'm responding to a question.

Mr. **Weiss**. Yeah.

Ms. **Wigdor**. And I'm on the spot and you notice I hesitated.

Mr. **Weiss**. Yeah.

Ms. **Wigdor**. Nevertheless, I point out that there is a problem with the hopes of many that truth in testing is going to help the disadvantaged and the question you've asked about what——

Mr. **Weiss**. As a matter of fact, it may very well be that it's the disadvantaged minority kid who wants to get into Harvard or Yale who may in fact be the beneficiary of the additional 25 or 30 or 40 points which will make the difference as to whether he or she is cut off or not.

Ms. **Wigdor**. OK. let me answer your question a better way then. I am now thinking through the report and through the chapter and indeed the committee would have given you a different answer and that is, and perhaps this is one of the more important points of this discussion of college admissions. That is that no test score should be the sole factor in admissions decision.

Mr. **Weiss**. That's true, but in the basis of your research, have you in fact not found that a large number of colleges and universities use the test scores as a cut off?

Ms. **Wigdor**. Yes. And that is to be recommended against. Yes, that is true.

It's usually a very low cut off, nevertheless, we recommended against it's use.

Mr. **Goslin**. It depends on the school, obviously.

Ms. **Wigdor**. It depends on the school, yes. We do indeed recommend against the use of scores, cut off points, with these kinds of tests and some other kinds as well. Yes, indeed.

Mr. **Weiss**. Well, I thank you very, very much for your testimony. I look forward to reading the full text of the report and, perhaps, we will have you back another time when we have the benefit of the report itself.

Ms. **Wigdor**. Thank you, and I hope we can look forward to seeing you February 2.

Mr. **Weiss**. Well, I look forward to it.

Mr. **Goslin**. Thank you very much.

Mr. **Weiss**. Mr. Kaplan, you have been the personification of patience.

Mr. **Kaplan**. Inside, not so, not so true.

Mr. **Weiss**. Well, I very much appreciate your volunteering to come down, responding to our request to come down. Then patiently sitting through everyone else's testimony and, without objection, your entire statement will be entered into the record and you may proceed as you so desire. You may read it in it's entirety, excerpt from it, summarize it, whatever you think is most appropriate.

[The prepared statement of Stanley Kaplan follows:]

PREPARED STATEMENT OF STANLEY H. KAPLAN, FOUNDER, STANLEY H. KAPLAN EDUCATIONAL CENTER

Gentlemen:

I appreciate this opportunity to speak about H.R. 1662. Permit me to introduce myself. I am the founder of the Stanley H. Kaplan Educational Center and have dedicated more than 40 years to the development and review of academic skills and content in preparation for tests at the high school and college levels. I began my career as an educator long before most standardized tests were even a gleam in the testmaker's eyes.

First let me say that I support the basic premise of H.R. 1662 -- that the public (including citizens, educators, and students) should be made fully aware of the subject matter, uses and procedures in which standardized tests are developed and administered. I am pleased that H.R. 1662, while suggesting constructive changes in standardized test procedures, does not attack standardized tests themselves. Standardized tests do have their place in the admissions picture -- a very important place. They more directly test the skills and abilities that are needed by students at higher levels of education. One's grade point average is an insufficient criterion, by itself, to evaluate skills and abilities. Secondary schools and colleges have different standards, and grades are frequently inflated. Certainly standardized tests, together with the student's GPA, give a more accurate picture of the student's ability than does either measure independently.

Standardized tests are important in evaluating the skills of an increasing number of adults who are returning to school after many years. Standardized tests also serve as an excellent device for motivating students to sit down for serious study. When the Wirtz report was released showing the steady decline in SAT scores, enrollments in my classes showed a dramatic increase despite the decreasing importance of SAT scores in college admissions. Parents were concerned about their children's weak math and verbal skills which they believed might impede their children's success at college. Standardized tests also help to identify underachievers including disadvantaged students. My years of experience as an educator and as a test preparer have shown me that an economically and socially disadvantaged student scoring a 550 (on a 200 to 800 scale) has as much potential as a student born with a golden educational spoon in his mouth.

I have dealt with some of the benefits of standardized testing because there are many who would like to eliminate standardized admissions tests, and view this bill as a first step toward the eventual elimination of all standardized testing as a means of evaluating students for postsecondary school admissions.

To repeat, the basic premise of H.R. 1662 is sound. I believe, however, that certain provisions in the bill are ill-advised and unworkable. One of these provisions, section

5, requires testing agencies to release to the test subject a copy of the test questions along with an answer key and a copy of the test subject's answers. This poses a number of problems. Test construction is a time-consuming and expensive process. It includes writing test items and then standardizing them so that the test subject can be properly evaluated. The nature of some tests restricts the number of test items that can be developed. Accordingly, valid tests cannot be produced on an assembly line basis. If tests are to be released after every administration, the number of test administrations must necessarily decline -- perhaps to zero. This is clearly borne out by events that have occurred since the New York State Education Act took effect. For example, the April LSAT exam was eliminated. The GRE exam in New York State will not be given next April or June, although it will be administered throughout the rest of the country. The LSAT and GMAT will be administered in June on a Wednesday evening. The AAMC has obtained an injunction against applying the disclosure provisions of the New York law to the MCAT. Small volume exams such as the VAT, OCAT, PCAT, DAT and College Board Achievement Tests were not administered in New York State until a series of amendments exempted these tests from test disclosure. The reasons are clear -- tests cannot be mass produced at will.

In view of these problems, why should we even have test disclosure? Proponents of test disclosure claim that there is a need to release test questions to students so that they will not have to pay large sums of money to coaching schools. There is no evidence at all to support this proposition. Students are not attending coaching schools to obtain test questions; they are looking for academic interaction with a teacher. Students are looking for an organized commitment to review materials they may have learned years ago, and to improve skills that were never fully developed. After all, questions of the type that appear on the tests have been available in review books for years.

Releasing test questions has, instead of leading to the demise of coaching schools, encouraged their growth -- with some schools even advertising they have the released questions.[*/] After all, when the Regents examinations were important in New York State, I had thousands of students preparing for these examinations even though the exams were released, and even though I myself wrote the test explanations in the Barron's Regents Series. The CPA examination is perhaps the most coached-for examination; yet the questions are released along with full explanations. As I had predicted in a letter to Governor Hugh Carey before the New York Standardized Testing Law was enacted, my enrollments have increased significantly for the released-tests -- LSAT, GMAT, GRE, and SAT -- but have remained at a plateau for other tests.

[*/] See attached advertisement from the New York Times (p. 8).

In addition to some of the problems I have with test disclosure, I am concerned that some of the statistical correlations the bill mandates may be difficult to develop. How can one make any correlation between test scores and success in a career? After all, how does one measure success? How can one draw conclusions about the extent to which test preparation courses improve test scores? There are literally thousands of programs of all kinds given in teacher's basements, in community centers teaching adult education, in high schools, and in colleges. It is impossible to identify the students being coached, or to develop a statistically valid means of measuring the amount of improvement because of coaching. Besides, even if conclusions could be drawn about the extent to which test preparation courses improve test scores, the "poorer" cram courses would be permitted to take a "free ride" on the benefits afforded by effective test preparation courses. All test preparation courses are not created equal. Even now, there are "quack" correspondence courses that are using the results of the FTC investigation concerning the SAT to promote other test preparation courses that were not even studied by the FTC.

Another problem I have with H.R. 1662 is that the disclosures introduce a number of inequities. Requiring the testing agency to release information has proven to be extraordinarily costly, and has already been reflected in the prices paid by the students to take standardized tests.

It is unfair to expect the economically disadvantaged student to pay these increased costs. I also believe it is unfair to require disclosure on high volume tests while exempting low volume tests from disclosure. If it is so important to see the test questions, why deny the student who is taking a low volume test this "important" opportunity?

So far my testimony has offered only criticisms of various provisions in the proposed legislation. I do, however, have some suggestions. In section 4(a)(b), for example, it would be more meaningful to determine the relationship between the grade point average of students and their income, ethnic, sex, and handicapped status rather than the relationship between these demographic factors and a standardized test score. After all, it is the GPA, and the achievement level through high school and college that is still the significant factor in admissions. A standardized test is simply a mirror of academic achievement in school.

With regard to the release of examination questions, I have found in my talks with thousands of students that their chief concern revolved around the correct scoring of their examinations. Their interest in actually seeing and going over the questions was minimal. After all, they had already taken the test. I have at times offered to go over questions students may have had about their released tests. But there have been very few takers. My suggestion is that testing organizations be required to send each test subject a copy of that subjects answer grid and a copy of the corrected answers so the student can check the scoring. I do not believe that tests should be released after each administration, but instead once in every four administrations and to every student before he or she takes the test. Releasing tests once in every four administrations would lessen the pressures on the testing agencies, allow the tests to be given more frequently, and still give consumer groups and educators an opportunity to judge the accuracy, ambiguity and possible cultural bias of the questions.

Finally, I would like to say bravo to section 2(a)(2) of the proposed legislation. I agree that "there is a continuous need to insure equal access for all Americans to educational opportunities of higher quality." The testing process, however, has only minor relevance to the equal access of educational opportunity. When we talk about equal access we should emphasize assistance to deteriorating schools and to disadvantaged students who just are plain not learning. In fact, there are many who aren't even aware that equal educational opportunities are available. We must endeavor to provide adequate education to all, so that everyone can fulfill his full potential. This is the area in which we should concentrate our energies. This is where the consumer -- the student -- is being shortchanged. Given equal educational opportunities there should be fewer inequities as far as testing is concerned.

THE NEW YORK TIMES

COLLEGE BOARDS-SAT

BARRON'S
TEST PREPARATION CENTER
35 HILLEL PL.
BROOKLYN, N.Y. 11210

CALL DAY OR EVENINGS
434-5242
444-7346

OUR 1 DAY 7 HOUR PROGRAM IS DESIGNED TO SIGNIFICANTLY BOOST YOUR S A T SCORE THIS FALL

- Designed specially for the Oct. 10th exam
- Practice and learn on actual exams
- Taught by N.Y.C. licensed teachers of Math & English
- The only program of its kind in N.Y.
- Offered weekends or weekday evenings beginning Oct. 3, 1981
- All materials included for only $45.

BE PREPARED . . .
IT'S PROBABLY THE MOST IMPORTANT
TEST OF YOUR LIFE
BARRON'S . . . the people who know

STATEMENT OF STANLEY H. KAPLAN, FOUNDER, STANLEY H. KAPLAN EDUCATIONAL CENTER

Mr. **Kaplan**. Well, I appreciate being last because one learns a lot from the previous speakers and, therefore, I sort of scrapped my whole official presentation that we submitted, that I submitted, and from there I have taken some notes. It follows along the same lines, but I have learned a lot from this and I hope that you may find valuable some of the——

Mr. **Weiss**. And there's no fee involved?

Mr. **Kaplan**. So I appreciate—I, myself, decided to come here. When I was at the National Education Association convention to which I was invited, in Michigan, to speak on how to prepare for the SAT. The NEA is quite opposed, in general, to standardized testing or at least very much in favor of the release of test questions, but Congressperson Chisholm, in a speech, she was talking about this truth-in-testing if you want to call it that. I like to call it standardized testing or something like that because I think the former is a sort of pejorative appellation, but she said if we have this bill passed, then students will no longer have to pay $300 to go to coaching schools and since my fee is about $300, I figured she was talking about my program. That's the reason why I felt I'd like to present some of my opinions, some of my views, because I have had the most experience directly, I hear this talk about coaching, the effects and I experience this with students.

I'd like to tell a little bit about myself. I'm the founder of the Stanley H. Kaplan Educational Centers. I've always loved to teach, from high school on when I was paid 25 cents an hour as a tutor, and believe me, I have been coaching long before a standardized test was a gleam in a testmaker's eye. These tests weren't around and I don't think that coaching should be tainted. It's not a tainted word. I think there are many, many beneficial effects of coaching. Coaching existed long before the SAT or the others came about.

I today meet former students who are professionals now, doctors, lawyers, who might have dropped out of high school, might have dropped out of college if they didn't have the support at the time. That's why when you have coaching, it's not unfair, coaching is remedial work. It's a kind of supportive work. It has nothing to do with admissions testing itself. It will always be here. What will you do if the standardized testing act goes through? You'll be finished. I say there will always be a need for coaching, for preparation for tests, for remediation.

I'd like to say first of all that the basic premises of this H.R. 6612, is it? Is that the right number? 1662, sorry.

Mr. **Weiss**. We change the numbers every few years anyhow, so it's all right.

Mr. **Kaplan**. The premises are excellent and I'm in support of many, many features of this bill. But, obviously, from my being here, there are certain features I object to, from experience.

I believe it's most important that the test subject be made aware of the uses and procedures of the test. Even facts about coaching. Remove the mystery of the test. I think that's great because there has been a lot of misinformation that has been going around.

I have a copy of a College Board bulletin from the—I must have it here somewhere. Yes, from the 1950's and this is from the College Board on the SAT. "If you feel you need coaching or tutoring aid, get it well in advance of your test date. Seek the aid of good teachers who will help you to analyze, to read with understanding, to reason through problems, to look for implications," and so forth. "The result of attempting to improve your scores by practicing with questions is not worth the effort. And to this philosophy I totally agree and I was doing this for many, many years.

Then in the 1960's, apparently the position of the College Board changed. It felt that coaching didn't help and when I was on a National Public Radio program in the 1970's, 1977, I believe, Dr. Knowland from the College Board, I believe, and I asked him, why the change? The test hadn't changed. Why did the position of the College Board change? And he said, and I can understand it, because the of proliferation of cram schools and which weren't having much of an effect. The result was that the College Board felt that they should not endorse these kind of programs which would cause the proliferation of the coaching schools, of cram courses.

Then, as I'm sure you're well aware, the position of the College Board has gradually changed in the late 1970's and the 1980's until now, in the bulletin entitled "Taking the SAT," which is issued to every student, point No. 3 states:

> Development of these abilities—the abilities tested by the SAT, is related to the time and effort spent. Short-term drills and cramming are likely to have little effect. Longer term preparation that develops skills and abilities can have greater effect.

This is the kind of thing that I would like to see disclosed because it certainly makes a separation between cram courses and there has been a lot of discussion about this, and long-range coaching and I think that's important.

Also I think it's important that the meaning of the word "aptitude" be clarified. It's been bandied about quite a bit today. The American Heritage Dictionary defines aptitude as inate abilities or developed abilities. They've got both of them and they're both in the same definition.

I have suggested many times that the A in SAT be changed from aptitude to assessment, because that's what the SAT is really doing. It's assessing your abilities. It's not pinpointing you as being a certain kind of student as an IQ does and that you'll be at that level for the rest of your lives, but I found out, in my talks to parents, in radio and talk shows, that many, many parents and students are concerned that they have been listed or blacklisted as a 550, 400, or 700, that's not a blacklisting, of course. The point is I don't believe that the SAT should be a label and the ETS has gone along with that. I've had many, many students taking the LSAT and the GMAT, the GRF, these are the same kind of tests; test the same kind of skills as the SAT; they come to me. "I did very poorly on the SAT. What does this mean? Does this mean I won't do well?" Again they have that label and I've said that it's only an assessment. My reply is, "You might have been a little pipsqueak in high school, never read, weren't motivated, watched television all the time." Suddenly you matured and I say, most of you, I guarantee if you took the SAT again your score would be much higher because you have improved. An improved student can get a better score. If you don't improve, if your skills aren't stronger, then you won't.

OK, I was saying I felt that the basic premises of the H.R. 1662 are excellent. We need more openness. However, I'm very happy to see that this bill did not attack standardized tests per se, because tests do have values.

First of all, the tests are goal oriented. In other words, they test skills that are specific and this is why I am talking, for example, to many, many deans of admission, members of admission committees at various professional schools. They've said that these tests are a better prediction of success in the first year of law school or medical school, than the grade point average. After all, the professional schools go to the testmakers and tell them the kind of skills they need tested, the kind of skills they feel will be most helpful in the professional schools. The testmakers then go about and make a test to test these skills and I know that many of my students have reported back to me when they were in college, when they were in law school or medical school that the kind of skills that were developed in my programs were very helpful later on. So I'm glad to see that my kind of programs are not just for the tests, but for improving the skills of students to do better later on.

So this is one value of the tests. They are goal-oriented and, therefore, can provide some more accurate assessment of the student in reference to where he is headed.

Also, there is the problem of grade inflation. Marks can be way off the wall. I know there are certain colleges in California where the average grade is about an A–. I attended a workshop at an AMC convention where there was a discussion on admitting high school students to 6-year medical programs and they were talking about the criteria, the criteria they needed for these 6-year medical programs and everybody mentioned the ACT or the SAT and my question was, "What about the grade point average? That's the most important criterion" and they said, "Well, everybody comes to us with perfect averages and there are still three times as many applying." They come with the 4.0's and there are still three times as many applying as we can accept so what can we do? So, therefore, we use the SAT for determination for a decision.

Of course, there are other factors too, but this test certainly, together with the GPA, can form a more reliable basis for decision than either one alone. Other values of the tests? There are adults, many more adults, today taking these examinations. I would say the median age of the students in my schools is over 25. There are parents, mothers who want to go back to professional work. Adults want to change their careers. Now you can tell me, how accurate can a record, an academic GPA be, when the academic record may be 15 or 20 years old? Therefore, this test again, can give some idea of where the adult is at the present time.

Also, I believe tests are a very motivational device for learning. I know, I've heard many times from parents and students: Mommy says read. The child watches television. Then Mommy says, "If I suggest that reading because it will be on the SAT, all of a sudden they start reading."

The Wirtz report which was issued, I believe, in 1978, told about the decline of SAT scores. With that report came an immediate increase in enrollment of my students, a tremendous increase. I was wondering why until I realized that parents weren't interested in SAT scores because it was relatively easy to get into most, except the most competitive schools. It's relatively easy to get into college. The SAT scores are not that important. Colleges are looking for students. But what the parents were concerned about was the lack, the weakness of the math and verbal skills as shown by the Wirtz report and, therefore, they were worrying about how their children would do at college, not about a SAT score.

Also, I believe these tests can identify under-achievers. There are many people who don't work at their real potential. They are bored at high school. I'm talking about the SAT now. And, therefore, their grade point average may be below what it really should be. I know the SAT has provided many opportunities, when the student did well—to allow the student to proceed on to a college. When I say under achievers, I'd like to address the fact of the disadvantaged student because many, many disadvantaged students may have great grade point averages, but so many colleges distrust the grade point averages because of the inflation and when the student does better or does reasonably well, then the SAT is an excellent persuader as far as the college admissions committees are concerned.

I would say, very strongly, that if a disadvantaged student from a disadvantaged background can achieve say a 500, 550 on the SAT or on the LSAT, even below a 500 on the LSAT, then that student has as much potential as an advantaged student who achieved a 700 and was also born with an educational golden spoon in his mouth, his or her mouth.

Also, some mention was made of the test validity and I've spoken to many people, many admissions committee members and I would pose this question, "Who are the people to go to to find whether the tests are valid?" You go to admissions committees. The admission committees are under no compulsion to use the results of the tests, if the tests didn't work out to their satisfaction. They can immediately dispense with these tests.

I do know, for example, that I—this is just an isolated case--I spoke to a dean of a dental school down South. The school was having trouble getting students, so they decided to lower the standards for the DAT and most of those, who had the lower standards, flunked out after the first year. So apparently these tests do have some validity, but of course, I'm not the expert in terms of making scientific investigations.

The reason why I have been talking about the values of these tests is because there are many people who argue against the tests, who would like to see the tests eliminated, and I believe that these same people are supporting the test disclosure provision of this bill. Hopefully, they expect that the weaknesses or the increased unreliability of the tests might eventually lead to what they hope will happen, namely, the elimination of standardized admissions tests.

I do know, I spoke to a member of the admissions committee at a medical school—I know the AAMC is much against the release of the tests as you mentioned before—and this admissions committee member said that if the MCAT isn't given, then we'll go back to the students that are coming from the Ivy League school types because their grade point averages are much more credible and of course, then we're back to the situation of the 1930's and the early 1940's where if you went to elitist schools you had a much better chance of getting into elitist colleges and elitist professional schools.

Now as I said, there are many, many helpful provisions and valuable provisions in the bill, but I believe there are some that are unsound and even unworkable.

From my experience, the biggest concern that students have had is whether they were scored properly or improperly. I think that's important. But releasing the questions of the test, I think creates problems, and actually, I think it's almost a tempest in a teapot. Because I've spoken to many of my students after the test—I've even offered to have them bring in their released test and we'll go over it with them. They showed no interest. After all, they've taken the test already.

Of course, one might say, they might take it over again. But most are not interested. And for example, if you take the LSAT again, then for most law schools, they take the average of both scores. And that's not so good. And it would be a much better idea—and I'll approach that in a moment—to release tests before they take the examination. I don't mean the test they're going to take, but others.

But I still think it's a very, very difficult procedure in terms of making these tests. I know I, myself, in making up my tests—I have a large staff. I make up materials for the test, and I don't have to standardize them. And it's still a tremendous, tremendous problem.

And the result, of course, is the number of test administrations decline, and in some cases I'm afraid if it's implemented, the number of test administrations for certain exams may decline to zero. Now, what's some evidence of this?

The April LSAT has been eliminated. The GRE in New York State will not be given in April or June. The June tests, a very popular testing date for the LSAT and the GMAT test day; these tests will be given on a Wednesday evening. I don't know how eager you would be to take a test on a Wednesday evening after you've worked at a job all day. Of course you'll say, well, let them take it another time. But the June testing date has always been a very popular testing date.

The MCAT, as you mentioned, is not subject to test disclosure, because of an injunction in New York State. But I do know that the VAT and the OCAT and PCAT and the DAT and achievement tests were not given in New York State until they received exemptions by a series of amendments.

Now why did they do this? Why didn't they give the tests? Because I believe they felt it was almost impossible and incredibly expensive to produce these tests. And that's why they rather would not have given them in New York State.

Now of course, if it's made a national law and the testing organizations have to release the tests. I suggest they might only be given, perhaps, once a year if they find that's necessary. But I don't believe this is in the interest of the students, because if a student gets sick, then he has to wait another year. And that's not good.

Now why have test disclosure? I'd say one of the most important reasons for having test disclosure was the statement that students would no longer have to go to coaching schools where they get the questions from the tests.

Now the assumption was they get the questions from me, or from coaching schools. But they could also get the questions from textbooks for $5. I have a staff make up our own questions and I believe they're of a little better caliber than the ones from the textbooks, but they're available. But as far as students not needing coaching schools, the evidence has been—and this is the main point I wanted to get at—has been quite to the contrary.

I found from my experience that students aren't interested in questions. They're interested in learning. They're interested in an efficient way of organizing the study materials, reviewing the materials. They're interested in the interaction of a teacher, not just having questions. Now someone mentioned that the students will see the question and see where they went wrong. I doubt very much if the student will see why he or she went wrong, because there's just a key. (B) Why is it B? The student doesn't know. He knows he has that wrong, but he doesn't know why he or she had that wrong. And there are no explanations that are given as to where the mistake was made.

And as I mentioned before, the students aren't even interested, because I offered that class after they got the test back, to have a workshop to find out where they went wrong. And nobody showed up.

Now I mentioned all this in a letter to Governor Hugh Carey before the truth and testing law in New York State was enacted. I predicted that there would be a run to coaching schools, a rush to coaching schools, once the questions were available. And it turned out just as I predicted.

Now how did I know this? I wasn't any kind of seer into the future. I went by past experience. The regents examinations have always been released. And I myself was the author in 16 titles of the Barron's regents series, with full explanation. And when regents were important for admission to the New York State schools in the 1950's and 1960's, the students flooded my programs. I had thousands preparing for the regents examinations, even though all the questions were published with full explanations, published by me.

So obviously, there's more to it. They need a commitment. They need the interaction with teachers. Take the CPA; this is another reason I knew this would happen. The CPA is the most coached-for test that I know of in the country. Almost everybody takes the CPA course. And yet all the questions are released with all the answers explained completely, and yet everybody rushes to coaching schools.

So the result is, I'm glad to say, that enrollments in my programs for the released tests, the LSAT, the GMAT, the SAT—did I leave one out? GRE, I don't know if I said that—have greatly increased. But the ones that haven't been released, have stayed at a plateau. The DAT, the VAT, the MCAT—they've stayed about the same.

So you would think it would be to my interest to release all the questions. Unfortunately, I'm an educator, and I seriously think that the reliability of the test will be compromised and the students are being hurt, and I think the best approach would be some kind of compromise in terms of test disclosure, which I'll talk about in a moment.

So anyhow, the whole effect was counterproductive. In fact, I have—and I think it's in the released copy to you—a copy of an ad from a one day course by Barrons. But it's not the Barrons of the Barrons educational series. It's another, near Brooklyn College, another Barrons store which apparently is giving a program for the SAT. And here is, in a highlighted point, "Practice and learn on actual exams. Our-1 day 7-hour program is designed to significantly boost your SAT score this fall. The SAT is probably the most important test of your life. Barrons, the people who know."

Now, obvious, I'm not trying to denigrate any particular coaching program, but this is where I think the FTC really might come in handy. Because to say that in a 1-day 7-hour program one can really significantly improve SAT scores, I think is a farce. And I know that there are other correspondence courses, six lessons in math, and six lessons in verbal, $125 each; $250 for six lessons that's sent through the mail.

And they have the cheit—ah—let's use the word—to include in a comparison of their program with my program. They put, "Stanley H. Kaplan's program is this." And the problem is that of the 20 points they point about the Stanley H. Kaplan program, about "21" of them are wrong. So this is where I say the FTC can certainly come in, can certainly be very useful. So that's, I think, an allusion to the difference between cram courses and longer courses.

By the way, a lot of talk was made about the FTC. The FTC investigation started with me. I'm glad to say—or I'm sorry to say—I don't know which to say—because I stated in an ad in Boston "Prepare for the SAT," and the SAT, ETS—or College Board, I should say—said you really can't prepare for it. And obviously I sincerely felt I could. I've been doing it for 30 years.

So the result was an investigation. Now I still disagree with the results of even the Boston FTC report, which said, improvements could be made, and the FTC Boston region report, was not released. And when I show copies of it, I always have marked on it in very big letters, so I won't be considered having done things that were really not ethical. I have on the top of each page, "This is from the Boston regional office report, rejected by the FTC in Washington as seriously flawed."

But I still would go along with many of the points there in the Boston report that improvements in the SAT can go up to 50 points each. It means 100 points, which is about what I find. Although here in Washington, the report was much more diluted.

Now of course, these are the statistics that I've gathered. Now the college board has said, that my statistics come from students that are motivated. And I'll plead guilty to that. But nonmotivated students don't come to me. But most nonmotivated students really aren't interested in college.

Well, anyhow, with the FTC investigation,I really find that we need a lot more investigation in terms of what SAT improvement can be made—and this is important—in terms of seeing that false claims are not made by test preparation organizations Because I think there the consumer is endangered.

Now what are some of the inequities that I believe are implicit in this bill? Section 5. First of all, students will lose the opportunities to take tests, because as I have shown before, fewer tests are now being administered.

Also there are increased costs with the release of tests. It costs money to release tests and therefore, the students pay more. How of course, you'll say, well, the student doesn't have to request the test, and then he doesn't have to pay. Unfortunately this isn't true all the time. Because the LSAT since, I believe, the spring of 1981, now makes it mandatory for everybody to get his test back, whether the test taker wants it or not. And the cost is built in. It used to be $15, I believe. Now it's $20.

So therefore, even though the student doesn't want to see his test, he must still pay the additional $5. Also if the committee feels as the bill purports to say—that the release of test questions is very important, then what about those under 5,000 tests where the student won't see the questions? After all, you can't play it both ways. If it's important for the students to see the questions, and I know there are practical considerations involved, why should some people be punished, just because they're taking a test under a 5,000 administration.

Personally, as I said, I don't think they're being punished. I don't think most of them want to see their tests. So therefore, I don't think that's much of an argument.

Mr. **Weiss**. Your argument.

Mr. **Kaplan**. No, mine. Mine. I'm sorry.

Mr. **Weiss**. Your argument, OK.

Mr. **Kaplan**. I appreciate any clarification on your part. But what are my suggestions? One suggestion I think is very important, and I've been asking this for years and years, is that the student receive the official answer key and his answer key because most of the concerns of the students were whether they mismarked the grid. Was there a scoring error? Now for example, just the other day someone spoke to me about her friend's daughter who had failed the bar exam. And she was very concerned. She thought she had passed. The test is not released, but she went up to Albany and saw her test and found there was a clerical error, and she actually had passed with flying colors. Now obviously, for a student to see if he or she is scored correctly, is certainly, certainly necessary and only just.

Now how about students seeing a test before he or she takes the test? I think that's very important. But I don't believe every test has to be released because of that. My suggestion would be to release one out of every four administrations. And the student be given this test weeks before, not after the test. This way he'll have time to prepare. Now the students see the test only after the fact. In fact, although the college board says they began producing a former SAT before the New York State law, I think the fact that the New York State law was implemented and that there was more of a consciousness of what admission testing was all about, has really accelerated the pace of disclosure on the part of the college board.

And I would certainly say—in fact, I would like to see this written into law—that every high school be mandated to administer one of these tests in a test-taking situation for the junior class or the senior class, because just to give this to the student doesn't mean anything.

I hold courses for students who had taken the Junior SAT, and they didn't do that well. They are now with me to prepare for the test in the fall of the senior year, and usually when I ask how many have taken this released test, many of them had never even seen it, and most of them who had seen it didn't bother with it. So wherein is the value?

I think there has to be some intercession on the part of the school systems to really make sure that they know what the test is about. What the directions are about. How to pace, when to guess, when not to guess. This could be done by giving the test under the time restrictions with the pressures of an actual test, with people all around, and then going over the test and going over the kinds of pitfalls the students fell into.

This, I think, makes a lot of sense. Also I would suggest that when you're talking about the relationship of the admission test scores with income, sex, and so forth, I myself—it is my opinion—that it's the wrong emphasis. I think the GPA is a much more important factor to consider. After all, test scores are simply a reflection of what the student has learned in school. I certainly believe that if one reads the statistics you find that if you get a 750 you're in this economic status, if you get a 600, you're in this economic status, if you get a 400, you're in this economic status. However, if one looked at the grade point average, uninflated grade pointed averages, one would find the same thing. These are beautiful statistics about the relationship of scores and dollars. But as far as I'm concerned, it's a red herring. I've worked with disadvantaged students, and I see there's a direct relation between achievement, economic level, and parental supervision.

Now a lot of mention has been made about unfairness. Of course, all through life there is quite a bit of unfairness, and I agree. However, the parents always want the best for their children, and millions are being sent to private schools because they're getting a better education. They're getting better math and verbal skills so they'll do better on the SAT.

Does anyone think that we should have legislation abolishing private schools to remove this inequity and this unfairness? You mentioned, I believe, somebody from East Harlem. I believe his name was Patrick Shields, a very knowledgeable gentleman. He has these questions available. But I don't know, Representative Weiss, if you're aware of the fact that he contacted me and we gave him a program at no charge. A program in preparation for the SAT. Of course, he called me very late and it was a very short course. But at least they received some kind of input as far as the test-taking skills and as far as basic points which we felt would be very important.

In Texas, at the Health Science Center in San Antonio, they received—this is the third year now—they received Federal funding via HHS, to take about 50 disadvantaged students who had pretty good grade point averages—reasonably good—but poor MCAT scores. They were brought to San Antonio, and they were paid $100 a week, and they worked at the hospital, to see if this fitted in with their plans. They also worked with us on the MCAT's, and they have to attend.

For 2 years, the results have been pretty good, and from what I understand, their performance in medical school is also pretty good.

See, the big problem is opportunities. This is where I say bravo, bravo, bravo, to one of the points of the bill and to one of the purposes of the bill that there is a continuous need to assure equal access to all Americans—to insure equal access to educational opportunities of higher quality. But test disclosure is not going to do this. There is one way of getting equal access and that's to expand the opportunities, expand outreach into schools.

I go to medical schools, and they ask me do I know any minority members who might apply. They're looking for them. But many minority students absolutely have no kind of supportive work, no advice, no ideas of the opportunities. The minorities don't have this image of themselves as going into professions. The result is they don't train in science, for example, at the high school level.

And if some do go to the colleges, they flunk out the first year. If you wish evidence to that, I met a Dr. William Roth, Will Roth, at the AAMC Convention and he was trying to work with some minority students and found it didn't work because you can't start from scratch at the college level. The college level is not the point where remediation starts. It should start much, much before.

I believe this Texas experiment is working out very well with Federal funding. So I would think it's very important that the Federal Government help.

In other words, if test preparation helps, I believe the challenge is to see that everyone has the opportunity to be as well prepared as possible in schools that have adequate facilities and have skilled, dedicated and well-paid teachers.

Thank you very much.

Mr. WEISS. Mr. Kaplan, thank you very much. The bells have just rung. There's a vote on the floor. And I would like, if it's convenient for you, to take a recess for about 45 minutes and then come back and engage in some dialog with you. If your schedule is too tight, then we'll forgo that opportunity.

But if your schedule allows, that would be very worthwhile. What plane are you catching?

Mr. KAPLAN. I'll have to see. I'll try to make—I had an appointment to meet someone, a very important appointment.

Mr. WEISS. What time?

Mr. KAPLAN. But this is more important—2 o'clock.

Mr. WEISS. Oh, we'll do that.

Mr. **Kaplan**. Because I would—to me, I was looking forward to the other way around, your questions. Because that's when I learn. I don't learn from what I say.

Mr. **Weiss**. OK. Then supposing we come back here at—how about 1:30? Is that all right? Then perhaps we can let you go about 2 o'clock and maybe you'll catch a 2:30 or 3 o'clock shuttle——

Mr. **Kaplan**. I can reach this other—no, the appointment's here in——

Mr. **Weiss**. In Washington?

Mr. **Kaplan**. Yes.

Mr. **Weiss**. You'll do it?

Mr. **Kaplan**. I'd be here all day tomorrow if you want to speak to me.

[Recess for 45 minutes.]

Mr. **Weiss**.The Subcommittee on Elementary and Secondary and Vocational Education and post Secondary Education will come to order, and we will resume our hearings to H.R. 1662. Mr. Kaplan has been good enough to adjust his schedule so that he could be with us again this afternoon, will conclude his presentation, and then we may have a few questions to ask of him. Mr. Kaplan.

Mr. **Kaplan**. There was one item that I hadn't referred to and that was in terms of the impossibility of really finding out effects of coaching schools, because there are so many different types of coaching schools.

But also I am afraid that sometimes when coaching is approved on one level, then everybody gets into the act. And cram schools which perform nothing at all will ride on the coat-tails of those whose performances have really upped the level of the kind of improvement you can get.

For example, I know there's an outfit in California that advertises the FTC has now said that coaching helps. And they're preparing for the MCAT. But the FTC report in Boston judged only the LSAT and SAT and the released report only the SAT. People are generalizing, well, if you can prepare for the SAT, now take my program on the MCAT. Because the FTC has given the approval to all kinds of coaching, which, of course, is a very—it's a very dangerous thing to say, and I think it serves no purpose and can cause a lot of damage, because not all coaching courses are created equal. There are good ones and there are bad ones, and there are quacks and I think the FTC, as I said before, can do a good job in that area.

Mr. **Weiss**. Well, now tell me how we could deal with that problem. How can you really prevent preparatory courses or coaching schools which don't meet your standards from entering the field?

Mr. **Kaplan**. Well, I couldn't do anything. But I think the FTC could put up a series of guidelines, in terms of advertising claims. Because they send me up the wall. We promise you everything——

Mr. **Weiss**. Well, they may send you up the wall. I don't think that that's one of the criteria for the FTC getting into the act though.

Mr. **Kaplan**. Well, that's it. Because one cannot claim just because one student received a 750—obviously the implication is that you take my course, and you'll get a 750. In other words, there has to be—you can't take a few and advertise it—of course, it's misleading as far as the consumer is concerned.

All I'm asking is that there should be—talk about truth in testing. Truth in advertising.

Mr. **Weiss**. Right.

Mr. **Kaplan**. Don't mislead. The most important test of your life—it isn't the most important test of your life, and you shouldn't seduce people into taking the course, because of a rash statement like that which is completely false.

Mr. **Weiss**. You know that the new chairman of the FTC doesn't see very much justification for the FTC getting into the false advertising area. You've read about that?

Mr. **Kaplan**. Yes, I've read about that. To my regret.

Mr. **Weiss**. OK. Go ahead.

Mr. **Kaplan**. I'm looking for another letter that I wanted to read, and then I'll be happy, because I think it's a most important—yes, here it is.

Because there seemed to be a lot of focus on the unfairness. That some people can afford and some people cannot afford these programs. And I just wanted to read a letter of which I think tells a lot in terms of what coaching courses can do, and what a test can do.

This is a letter from the Valley Opportunity Council near Springfield, Mass. We gave a program for the SAT during the summer a year ago, 1980. It's dated October 1, 1980. I just wanted to read it rather than just introduce it into the record, because I think what they say can be very important:

> We at the Health Center Careers for Minority Students Projects support your efforts to find ways to increase the participation of low income and minority students in programs which provide students with test-taking skills and review of fundamental concepts of math and English. It has been our experience that exposure to this kind of a program is yet another avenue by which the educationally unaware student can begin to understand the process of upward educational mobility.
>
> All too often these students whose families have no experience in post secondary education, find themselves unable to fulfill career dreams because they're unaware of the complexity of requirements which must be met before they can acquire the formal education mandated by today's highly technical occupations.
>
> The steps toward college which the middle and upper class students take for granted as they pass through high school, remain a mystery to the disadvantaged student. In fact, the middle class student and his or her family is becoming more sophisticated in maneuvering their way through the college entrance maze.
>
> The disadvantaged student is expressing no comparable enlightenment. Like a soccer player suddenly playing football, he can't possibly score until he learns the rules.

And this is the important part:

> Test preparation programs provide a unique opportunity for disadvantaged students to immerse themselves in an environment whose goal is college entrance. They begin to see themselves in a new way. They begin to understand the challenge and how they must respond. We recognize that our project is only a beginning and are anxious to involve as many other resources as the student requires. The test-taking skill development is certainly one of these resources.

And that's about it.

Mr. **Weiss**. Right. Who wrote that statement?

Mr. **Kaplan**. That was written by a Jane B. Baatz, B-a-a-t-z. She is the coordinator for this opportunity council.

Mr. **Weiss**. I see.

Mr. **Kaplan**. Springfield, Mass.

Mr. **Weiss**. Good. Thank you very much, Mr. Kaplan. I really appreciate your testimony. I'm going to ask you some basic questions.

Mr. **Kaplan**. Right.

Mr. **Weiss**. And if you will, try to be as direct in your responses as you can. Sometimes the questions may not lend themselves to being—to giving brief responses. But I will try to phrase them in such a way that you can.

Tell me how your SAT courses are structured. We have in the record some indication as to how your preparatory course works.

Mr. **Kaplan**. Right. Well, it's basically the same as the other courses. What we have is, for the SAT, is 11 4- to 5-hour class sessions held once a week for about 10 or 11 weeks. And in that, we give the basic concepts. We review the basics of arithmetic—because they've had it before. It's a matter of review. We don't teach. We review. We go over the basics of arithmetic, algebra, geometry. They might have had it, but forgotten. Or never really understood.

And each time we test them to make sure, one, they've understood the concepts that we've given them, plus two, they're getting used to some of the test-taking techniques and pressures.

But then I compare this to a tennis coach who gives 10 half hour lessons. And what does the tennis coach say? He says, "Go out and practice."

And we have permanent centers. I think that's a very essential part of any kind of——

Mr. **Weiss**. What kind of centers?

Mr. **Kaplan**. Permanent. Permanent centers. In other words, they're not just hotel rooms, where you rent hotel rooms and then disappear.

Mr. **Weiss**. I see. Right.

Mr. **Kaplan**. We have permanent centers with facilities, where the students can work. And we have, what I call, our test-n-tape center. The test-and-tape approach where they can, first of all, review the lessons they've had in class, because sometimes they're afraid to ask the teacher a question. Sometimes they want to hear it at their own pace.

But then we have it all organized—and this, to me, is the crux of it—based upon the concepts we've taught in class, we have a whole series of tests that we made up. But we use the Socratic method, not just give an answer to the test. Each test introduces concepts and goes over the whole procedure on which the concept is based, and the procedure by which you analyze and think it out and get to the answer.

So it's not just a question and answer. Sometimes the discussion may take 15 minutes just on one question. That's why our cassettes sometimes are rather lengthy. But in this way, for example, in math on the SAT, you can take 1 question and make up what I call 10 different questions, exactly the same except entirely different. If you understand the concept, you learn how to think a little bit. And I'm not trying to make this too lengthy, but it reminds me of a student, and I think that's very important what she said. She said, "You know, Mr. Kaplan, I never thought of thinking." This happens all the time. Because too much work at the high school, the elementary school, unfortunately, even the college, is a matter of memory. And these tests get you off that kick of memory, and get you starting to be able to relate to concepts and to reason things out.

So we have about 250 to 300 additional hours. But there are menus and if students feel weak in certain areas, they can select those areas in which they feel weak. And this is the basic approach. Students can begin in January—most do—of their junior year. They take the test, say, in May or perhaps March. And in that time, they've had the time to go over the basic lessons and also to attend our test and tape centers. But then they can continue on. Past that test—through the summer, many of them take it again in the fall. So they continue on. In fact, many take it beyond the test date. In fact, I met—and this was a most vitalizing experience to me. I was in Columbus visiting one of my centers, and I met a youngster who was applying to a parochial school in Ohio. All he needed was about a 550, and on his PSAT's he had scored that already. But he felt there were some things to learn. And he scored a 750 and 780 when he took the test in May. Now, I was there in June. He was there. I said, "Why are you here?" He said, "I feel I'm still learning."

And that's the whole basis upon what my organization is based. In other words, not for the test. Of course, they're coming for the test. But to improve the math and verbal and reasoning skills. And that is a synopsis of what my program is like.

Mr. WEISS. How about the lessons in how to take a test itself?

Mr. KAPLAN. Well, we don't have lessons in how to take the test itself, per se. We're constantly, as I say, we give tests in class. We're constantly going over, pacing them, telling them when to guess, when not to guess. And so it's constantly gone over, in fact, not just in one specific lesson.

Now, of course, for very, very cram courses, that's one thing that can be done. As I mentioned before in my testimony, that's something that schools can do, is to give a sample test and go over it. And explain the procedures. I think that will be extremely helpful.

Mr. WEISS. Now how long have you been giving your courses?

Mr. KAPLAN. For the SAT?

Mr. WEISS. Right.

Mr. KAPLAN. Because I've been tutoring and coaching since 1938, for the SAT. In fact, you know I had never heard of these tests. A student came to me, someone I was coaching for the school subjects, and said, "Look. We have to take an examination of the SAT." And I said, "What's the SAT?"

And when they showed me the booklet, I realized it was all the math and verbal skills I'd been teaching all along. I said, "Well, I can prepare you for this." And so then it began to mushroom—it was the same test for all—so instead of teaching individually, you could have groups of students.

And word of mouth—in fact, I didn't advertise until 1970. The FTC came to me in 1974. Before that I didn't. But when other cram courses came about which I felt needed an addressing to, that's when I began to advertise to show the students had a choice in terms of selecting various coaching courses, if they wanted to.

Mr. WEISS. And tell me again, it is your conclusion that the difference in score, between somebody who takes your test, takes your course, and those who do not, is what? About 50 points? Is that what you said?

Mr. **KAPLAN**. The average improvement is 50 points in the math and 50 points in the verbal. And interestingly enough, many times the verbal improvement is more than the math. You know, talk about cultural bias, one would think that the verbal would be more culturally biased than the math is.

And yet from what I understand, statistics from ETS show that, even though both scores are low, the math is lower on a relative scale than the verbal. And how can there be a cultural bias as far as the math is concerned? That's hard for me to see.

Mr. **WEISS**. And have you published studies in this field as to the improvement that——

Mr. **KAPLAN**. The FTC. I don't publish the studies. No. But I—in fact we have about a—I would say at least about a 90-percent response on the SAT. It's very difficult for the LSAT and for the GRE, because students are mobile. But here, they're at home. We can talk to the parents.

And we get about a 90-percent response. And our average improvement has been about 100 points. And believe it, I've heard of some coaching courses keeping records and the poor scores are thrown in the wastepaper basket. The good scores, they keep. I've heard this from advisers in terms of evaluation of programs. And our improvement is 100 points total.

Mr. **WEISS**. Is that—and has that been fairly consistent over the years?

Mr. **KAPLAN**. Fairly consistent. Well, I wouldn't say that. It's been going down. I would say in the 1960's the improvement was perhaps 175 points altogether. That's because students were more motivated to work.

I used to have class sessions, something like 32 class sessions. That means 32 times 4—128 hours. And therefore, you could have a much more intensive review if it was all in class. But the students aren't willing to come to class that much. The SAT's aren't that important.

You know, for the LSAT, the median score is about a 530, above the average. For the SAT, the median score in the math is about a 420, and the verbal about a 460. And—or maybe the other way around. But the point is—one of the reasons I think is that people aren't paying much attention to the SAT. And therefore, if you don't work at it, you're not going to improve.

Mr. **WEISS**. And you know, there have been suggestions over the course of the years that courses, such as yours, in fact have gotten some of the questions from the kids who've taken the examinations. Do you want to address yourself to that at all? As to whether you utilize any of the test questions that have been given by debriefing the kids who have taken them, or anything of that——

Mr. **KAPLAN**. Let me say this. Debriefing doesn't help. We get an idea of the kinds of concepts that are given. But I know this was in the Boston FTC report, and there was one organization that boasted—in fact, whenever they went to talk to pre-law societies, that they regularly sent in people to take these examinations to get the questions.

And on the LSAT, the FTC report on the LSAT—this is the Boston report. They were at the bottom of the list in terms of effectiveness. If people come in just to memorize questions, you've got problems. And my course is not based upon having the questions. I don't see any kind of result, unless you have the exact question. But usually it's more trouble than—because people cannot, when they go in, they cannot remember the questions per se, unless they steal the test. The result is their impressions, like having an impression of elephant. One is the tail, and one is the—whatever. The impressions of the questions are entirely different.

It is very, very difficult to memorize questions just as they are. So therefore, when a student sees a question that is very much the same as it was on the test, ordinarily the student says, "Well, I don't have to read the question. There's the answer." And that's trouble.

Some students have come to me and said, "You know, Mr. Kaplan, the exam was exactly the way—you had all the questions." These students are going to get the 450. But the students who say, "Mr. Kaplan, you know, it was altogether different." These are the people who get the 600. Because what they thought was the same, really wasn't the same. The former had learned to memorize, not to think.

And I tell all my students when they come in at the first lesson, "If you're coming here to think you're going to get the questions, please get your money back now. We teach you to think, to understand, to learn, and then you'll get the material."

But I do admit that if there are any new concepts that are given on the tests and the students come to us—we don't ask them, debrief them, "What are the questions? What problems did you have?"

And if there are problems, we take note of them, and therefore introduce them into our programs. That's not the same thing. I'm dead set against questions just being memorized and given in the course, because it's a fraud as far as the student is concerned to think that he's going to be able to do that well because of it.

Mr. WEISS. Do you have any idea at all as to what percentage of the people who take the SAT examination, in fact, attend a formal preparatory course, such as your school or any of the other schools around?

Mr. KAPLAN. It is very difficult. That's extremely difficult to, because there are so many teachers in the basement of their home.

Mr. WEISS. Right. OK.

Mr. KAPLAN. They give courses, and schools, they give their own preparation. I know James Madison High School and Lincoln High School for years were giving coaching courses.

Mr. WEISS. Do you have any way of making any kind of intelligent estimate as to the percentage of test takers of the professional tests? LSAT or MCAT, who in fact, take preparatory courses?

Mr. KAPLAN. I would say that's much higher, because it's much more important. And I would say perhaps—well, perhaps the MCAT's, that's, you know, half.

Perhaps the LSAT—it's harder though, maybe 10 to 20 percent. I don't know. You see, even if you have a textbook, I mean, that's a coaching course.

Mr. Weiss. Right.

Mr. Kaplan. I mean you're working with the book.

Mr. Weiss. Right.

Mr. Kaplan. And some people can work with the book much better than they can work in class. So to estimate how many people use a review book, I don't know. But it must be a great number.

I would say a person would be pretty silly to go into the test cold. And if he's going to coach himself by means of review book, fine. But to take the test cold, I think is certainly a very inappropriate method of attack.

Mr. Weiss. We had testimony both at the beginning of these hearings back in 1979, and then yesterday from Dr. Cooper who represents the MCAT.

Mr. Kaplan. Yes.

Mr. Weiss. And you know, they just developed a new examination which they're very proud of.

Mr. Kaplan. It's not that new. Since 1977.

Mr. Weiss. Right. Well, relatively new.

Mr. Kaplan. OK.

Mr. Weiss. And they're very proud of it. And they take the position that because the test focuses on a particular area of the beginning college course, and a number of science areas, that if they were to disclose those examination questions and answers, that, in fact, they would run out of questions to ask.

And I wonder if you have any comment on that at all?

Mr. Kaplan. Well it's very difficult to give a definite answer but from my knowledge, and I was a major at science in college—in fact, I applied to medical school, but being the tender age of 17, I was turned down. Also my name was Kaplan, but that's another story. There is a bulletin, the MCAT bulletin of information, and it lists a series of all the topics in all the sciences, biology, chemistry, physics, math. And they list the topics they're going to ask. And they promise on a stack of Bibles it will be only these topics.

Therefore, they're narrowing themselves down to a very limited area. Of this is good information for the students, because the student knows exactly which areas to go over. In fact, the bulletin leaves out magnetism. Why, I don't know. But a student, when he sees that magnetism isn't included, right away says, "I won't study, review magnetism."

This bothers me. I teach it in my course anyhow. But I say, our job is to give you a science review. And so it won't be on the test. So what? You learn something. Maybe it'll help you in medical school. Because they do use magnetism principles in medical school, too.

Well, the point is that because the basis for items is very narrowed, I can see Dr. Cooper's point of view. But I cannot see Dr. Cooper's point of view of not, at least, releasing, say, a test every 2 years. The bulletin has a sample test, but it's not a full sample test. They have a small sample test of about—there were 300 questions on the test, and they have a small sample test of about 75 questions. Now the bigger one, they didn't give it the first year. But the next year, they released a bigger test of 203 questions.

And I asked once why not a whole test of 300 questions. They said because we don't have that many questions to release, so we can't even give one full test.

So the AAMC must really be convinced that they can't do it they can't even release enough questions for one complete test. Of course, these questions are standardized, and some are very, very sophisticated, and they might be very, very difficult to be able to put those questions together.

And I know as far as remembering questions of those kind of tests, forget it. Because if you try to remember them, you're giving the students more problems.

Mr. **Weiss**. What kind of experience have your students had with the MCAT?

Mr. **Kaplan**. I would say very good experience.

Mr. **Weiss**. OK.

Mr. **Kaplan**. Especially—the minorities have improved very significantly. They experimented in Texas—that was a definite scientific proposal, because here you had the students. They had taken the MCAT once. They took my program over the summer, about 200 hours. And they took the test again after. And so here you have a pretty good kind of evidence, and it did, for them, produce a significant improvement.

Mr. **Weiss**. Can you quantify it?

Mr. **Kaplan**. The improvement was something about, I would say, 1.2 points per subject. Something like that. Not as much as other students, because remember, these are disadvantaged students who have had poor backgrounds to start with.

But in other words, the percentiles would go up, say, from a perhaps 20 or 30 percentile to maybe a 50 or 60 percentile, in many cases. Some just from a 20 to a 40. Some didn't improve at all. But the fact that Dr. Medina, who is in charge of the program, asked for its continuation for another 2 years, and it was continued—this is the third time we've given it. I was just told yesterday that HHS had given them another—granted another proposal for the continuation of another 3 years, shows it must be some positive effect of the program.

Mr. **Weiss**. So that as far as your school and your courses are concerned, the fact that the MCAT's are not released at all beyond that one sample that they provide, is not a—it doesn't limit you at all in being able to instruct your students in the subject area of the test?

Mr. **Kaplan**. That's right. Because, in fact, I've never had more help from the AAMC, because they have that whole syllabus. Before I was floundering. There may be some botany. Now I know exactly which areas which will be covered by the test. I still expand on it. But they've been very helpful.

But you see, you still need review. Because in math, they say, go back and review the contents of a second year algebra textbook. Now when is the last time a student who's in his junior year in college been conversant with a second year high school algebra textbook?

And a good student can study on his own. But you see, interestingly enough, we have, for instance in Boston, students from the very prestigious schools, such as Harvard and MIT taking our courses. Why? Because everybody admits you must review. The AMC in its booklet says you must review. But students find that they need a commitment, because they have so many things to do. If they don't enroll in a course they really won't review. Their time is very, very, very constrained. And therefore, they are looking for opportunities of having an efficient means of review. And our program provides that.

Mr. WEISS. Mr. Kaplan, we have to break again. There's a vote on the floor. We'll be back in 10 minutes. The committee stands in recess.

[Recess for 30 minutes.]

Mr. WEISS. Mr. Kaplan, again, the committee will resume hearings and, again, my appreciation to you. We were talking about the MCAT and your capacity to teach that within the perimeters of what they test, even though they have not disclosed any of their examination questions except for the sample test that they provide.

Mr. KAPLAN. The sample test give a very good idea of what it's about. In fact, you know, the first exam that was given, we did very well with it even though it was an entirely new test. And the samples that they gave were an excellent guide for making up questions. We have a large staff that's continually making up new questions, based upon what we think are important.

Mr. WEISS. Have you had occasion from time to time to review the SAT questions and determine what percentage of them address specific subjects and what percentage of them really are reading comprehension?

Mr. KAPLAN. Are you talking about the math or the verbal?

Mr. WEISS. Verbal.

Mr. KAPLAN. The verbal? Well, as far as I'm concerned, the completions are a short form of reading comprehension. In fact the analgies are really another form of asking the vocabulary questions.

In fact the SAT, when they give their scores, gives a subscore in the vocabulary, and then subscore in reading comprehension. So even though there are four types of verbal questions, there are only two subscores. Do you follow what I mean?

Mr. WEISS. Yes. Yes, I do. Do indeed.

Mr. KAPLAN. So percentage wise, I mean, reading comprehension is reading comprehension. I don't understand what the alternatives——

Mr. WEISS. It's reading comprehension, rather than specific subject matter.

Mr. KAPLAN. Well, no subject matter whatsoever. You don't have to have any previous knowledge. In other words, the ACT sometimes gives questions in their reading classes based upon previous knowledge.

Mr. WEISS. Right.

Mr. KAPLAN. But here, you're told in the College Board directions, that all information is obtainable in the paragraphs that you're reading. Did I make that clear?

Mr. WEISS. Yes. Yes, you have.

Is there a requirement for there to be a previous knowledge of English grammar?

Mr. KAPLAN. Yes; definitely. Because there's one section, you know, the writing ability section, standard written English, and that is a definite requirement—you certainly have to know a verb and subject agreement. You have to know dangling participles and so forth. And I think that's very valuable.

Mr. **Weiss**. Now you have courses—you have classes that you give around the country now. Is that right?

Mr. **Kaplan**. That's correct.

Mr. **Weiss**. How many locations?

Mr. **Kaplan**. 105.

Mr. **Weiss**. 105 different locations. And I assume, but you tell me if I'm wrong, that you have different ethnic compositions in your student bodies, depending on where those locations are.

Mr. **Kaplan**. That's right.

Mr. **Weiss**. Have you been able to determine either through any kind of formal analysis or really by just some impressions that you have as to variations in performance of the students from various parts of the country on the SATs?

Mr. **Kaplan**. I don't know if it's a matter in different parts of the country. I do know it's when it is advantaged and disadvantaged. And that's in all parts of the country.

Mr. **Weiss**. Expand on that a little bit, if you will.

Mr. **Kaplan**. Yes. The disadvantaged students, for example, just give you an example in my own milieu of New York City. The students come in with a very limited knowledge of math. They're in their junior year of high school and can't handle their multiplication tables. They can't multiply fractions. They don't know what decimals are. And yet, they're supposed to take an SAT.

I'm not saying, I'm not using this in any derogatory fashion——

Mr. **Weiss**. Yes, right.

Mr. **Kaplan**. I'm just being very objective.

Mr. **Weiss**. Right.

Mr. **Kaplan**. In vocabulary, their vocabulary is almost nil. I'm not talking about street vocabulary. I'm talking—I don't mean that in a pejorative sense also, but the common vocabulary, fine. But if you take a word like "adamant", you mean stubborn, well, they have problems.

Mr. **Weiss**. What was the word? I'm sorry.

Mr. **Kaplan**. Adamant. Adamant.

Mr. **Weiss**. Adamant?

Mr. **Kaplan**. Yes. So that's a word one doesn't say to his friend, "I'm adamant." He'll say, "I'm stubborn."

Mr. **Weiss**. Right.

Mr. **Kaplan**. And so therefore, obviously if you're going to go to college, you have to have certainly a stronger vocabulary in order to be able to read and you have to have a lot of experience in terms of reading passage, how to get inferences from the passage, how to connect ideas, and find out the ideas that are being presented.

These are experiences that, unfortunately, the students haven't had—these disadvantaged students, because they've gotten a very, very inferior education. This was the point that—addressed in terms of equal opportunities.

The big, big problem is reaching out. I believe very much the way the Catholic Church believes. "Give me a child from 1 to 6, and I'll have him for the rest—him or her, for the rest of my life." There's a tremendous amount that has to be done in terms of a disciplined kind of education and a tremendous amount of assistance on the part of government to provide these opportunities.

But there are—still there are other minority groups who are not disadvantaged, and they perform just as well. I've had many—and I frequently ask—I see a black student who scored 700 on the Boards, on the SAT, and I say, "By the way, what do your folks do?" "Well, my father is a physician. My mother's a nurse." Whatever.

Family participation, family encouragement, is entremely important. I don't know if you've heard of the Beasley School in Chicago. That's a real demonstration. The Beasley School in Chicago where the superintendent—she was on "60 Minutes" once—said, "If God had believed in permissiveness, God would have issued to Moses the 10 suggestions." Which I think is very, very meaningful.

It's a disciplined, oriented kind of program. Students wear uniforms. The parents have to sign contracts with the school to go over the homework, to participate in the homework, and attend PTAs. And if they don't, the children are dropped. It's a very, very successful school.

The trouble is, it was $5 million federally funded project and it stops at the eighth grade. What happens from there on, I don't know. But to me, if you don't give the disadvantaged what they deserve, then we're going to get—as the advantaged—what we deserve in terms of the problems later on. And I couldn't feel more strongly about that.

Mr. **Weiss**. Have you been able to determine whether there is a difference in performance on either part of the SAT, based on geography—that is, what part of the country the test is given? Or by gender?

Mr. **Kaplan**. No. Neither by gender, nor race.

Except, of course, I'm saying, in terms of disadvantaged.

I haven't noticed any difference. Of course, I can't—I haven't examined all statistics throughout the country, but I visited all parts of the country, talked to my administrators, and I've asked this question a number of times.

And of course, you realize I'm going on the basis of a smaller number of statistics, because in different parts of the country, there's different amount of interest in terms of preparing for the SAT. This is a point that was brought out—or wasn't brought out as forcefully, that most people throughout the country aren't interested in the SAT on the ACT as far as the performance, because they're just used as sort of guides. You're accepted but the college look upon it—the scores—as guides to channel you once you're to college. But they're not used for admissions purposes.

In Ohio, if you have $400 in cash, you're accepted. So therefore, the SAT, since the standards are so low, except for the competitive colleges, the standards are very minimal and therefore the SAT score or need for coaching is insignificant.

Mr. **Weiss**. Do you find that the students who come to you for the most part go for the more competitive colleges?

Mr. **Kaplan**. I would say there are some that come to us just because of the math and verbal skills. But I would say, yes, more or less, they go for the—not the very competitive colleges. But I think part of it is that it's a combination of both.

The parents want to be assured that the children will perform at college at a reasonable level. And an SAT score which reads out the math and verbal skills, the parents feel coaching may be of help in giving this kind of reading.

Mr. **Weiss**. You don't really believe that the people who go to your courses go in order to have a refresher course in the subject matter, rather than to get a better grade on the SAT, do you?

Mr. **Kaplan**. They come for both. For both.

Mr. **Weiss**. Wouldn't you think that the overwhelming percentage, people who come to take your courses, take it in order to improve their chance of getting a higher grade in the course?

Mr. **Kaplan**. I would say so, yes.

Mr. **Weiss**. I would think so.

Mr. **Kaplan**. But I'm just saying that all the parents I've heard, say that their children have learned a lot. And therefore, even though they're looking for a better score, they realize—it's been said to me many, many times. "We're coming here to get a maximum SAT score. We also realize that our child will learn a lot based upon what they've heard from other students."

I couldn't advertise, "improve your math and verbal skills," and get many students. When I say, "prepare for the SAT," they all come. But whether they know it or not, they're getting a tremendous education as far as math and verbal skills are concerned. So that's the device I have to use. That's the device.

But I'm satisfied as an educator that they're learning something.

Mr. **Weiss**. That's doing well by doing good or something, right?

Mr. **Kaplan**. Something like that.

Mr. **Weiss**. Right. Listen, I really very much appreciate the testimony that you've given us. I guess one final question.

How did you feel during those years, which were until very recently, when the college entrance examination board people kept on saying, nothing you can do by way of preparation is going to help you. That you've either got it or you don't have it. And don't bother trying to prepare for these tests.

Mr. **Kaplan**. I kept saying, "This isn't so. This isn't so." And I spoke to advisers. This isn't so. In fact, this is interesting. A lawyer, after the "Phil Donahue Show" about let's get rid of the SAT's, a lawyer called me up from Washington and suggested I sue the ETS because of all the students I lost over the years who believed what ETS had said, that you can't prepare.

But of course, I'm not concerned about that. I'm just very happy about the fact that even the college board and ETS have acknowledged the fact that the SAT is not—well, I don't think they really meant it to be a test of innate aptitude. But they have acknowledged the fact that it is an achievement type of test, and therefore, that you can be helped. And a long-range program can help, while cram courses, most likely, won't.

And I certainly subscribe to that. But someone here asked me, don't you provide financial aid? Suppose you can't afford it? Because this is unfairness. And we try to go to all the advisers throughout the country, elementary, high school, and at college level and professional school level, letting them know that if a student is disadvantaged, and cannot afford the program. I don't mean reluctant to pay, but cannot afford to pay—if we get the documentation of the school, we will provide the service even though they can't pay.

Mr. **Weiss**. And isn't it also somewhat ironic that the FTC studies were precipitated by a question as to whether your claims were valid or not as to whether, in fact, you were able to help people prepare for those tests? And it ended up being an inquiry into whether the college entrance examination board, and ETS, was being accurate in its representation that in fact you couldn't prepare for it?

Mr. **Kaplan**. Sometimes, you know—this is interesting that I'm suspicious that that's what they—the FTC—had in their mind originally. That they knew that my work was, you know—that my programs were producing results. And this was sort of a indirect way at showing that the claims of the testing service weren't necessarily what they professed to be.

So I'm not sure——

Mr. **Weiss**. You're really an optimist at heart, aren't you?

Mr. **Kaplan**. Well, I do know that Arthur Levine resigned right after and he became the lawyer for NEA, National Educational Association. And he's a strong supporter, antitesting supporter. So again, I don't know what the original purposes were. But somewhere along the line, as the investigation went along, I felt they weren't targeting me. They were targeting the testing services.

Mr. **Weiss**. That's how it worked it.

Mr. **Kaplan**. But that's—well, I think, you know, it worked out at the end.

Mr. **Weiss**. Right.

Mr. **Kaplan**. But I think all along, I kept getting a feeling that that's—that was the target.

Mr. **Weiss**. You carry your shield of innocence. And apparently, it's worked very well.

Mr. **Kaplan**. Well, I don't think it's a matter of shield of innocence. I think it was a matter of confidence that knowing what I was doing, and I had full confidence in knowing the outcome. Because I couldn't have existed for 40 years. You can fool some of the people all the time and so forth——

But until 1970, I had never advertised at all. It was all by word of mouth. It was only after 1970, as I said, before this came about. So I don't think that more than 100,000 people over the years can be wrong. And it's a very satisfying thing as far as I'm concerned.

Mr. **Weiss**. Well, again, I'm grateful to you for coming down, for spending all of this time with us. And I'm sure that if we ever have need of calling on your knowledge and wisdom and expertise, that you will make yourself available to us.

Mr. **Kaplan**. Have satchel, will travel.

Mr. **Weiss**. Thank you very much.

I understand that Mr. Messick had to leave, but that Mr. Goodling has indicated that he will submit questions to Mr. Messick in writing, and those questions and answers will be then be included within the record.

The hearing now stands recessed, subject to call of the Chair.

[Whereupon the hearing was recessed subject to call of the Chair.]

[Material submitted for inclusion in the record follows:]

PREPARED STATEMENT OF ASSOCIATION OF AMERICAN PUBLISHERS, INC.

This statement is submitted on behalf of the Test Committee of the Association of American Publishers (AAP). AAP, which is the general association of book publishers in the United States, includes among its members companies from the standardized testing community. Several companies, however, while members of the AAP Test Committee, are not members of our association; namely, Harcourt Brace Jovanovich and Science Research Associates.

While several of our members provide tests for college entry, most of our members are the source of tests which are used in elementary and secondary schools or by private enterprise for hiring and promotion. Our concern, therefore, encompasses both the Weiss and Gibbons bills pending before this committee -- especially as they may be combined -- and our testimony consequently deals with each of these measures in some detail.

Before commenting on specific provisions of each bill, we should like to make some general observations.

Pending Legislation Discriminates Against Smaller Companies

The Commissioner of Education of the State of New York, Hon. Gordon Ambach, in his October 16, 1979 report stated that some 20 of 26 admission tests which had been available in New York in 1979 would be withdrawn from use in the State in 1980. It is interesting to note in this regard that the tests which have been the spotlighted targets of criticism -- tests from the Educational Testing Service (ETS) and American College Testing (ACT), the giants in the field -- would continue. But other tests, the products of smaller companies and administered in less-than-carload lots, were the ones being withdrawn.

While the New York State statute has since been amended so that the tests furnished by the smaller companies can survive, no similar effort has been made to alter the pending legislation. Much has been said at these hearings about the shortcomings of ETS and its Scholastic Aptitude Test (SAT) and how the pending legislation is needed to correct their alleged abuses, but the fact is that the Weiss bill, as experience with the 1979 New York statue after which it is modeled indicates, will permit ETS to continue its operations at the expense of its smaller, profit-making competitors. It brings to mind the old Mac Sennett silent comedies where the custard pie was aimed at one individual but hit another. It is the smaller companies which are getting the custard pie in the face.

This unconcern for the fate of the smaller tests was most recently pointed up in the November 4 testimony of Mary Ann Austen, Counsel to New York State Senator Kenneth P. LaValle, the sponsor of New York State's Truth-in-Testing Law. When asked about tests which were no longer being given in New York State, Mrs. Austen, who had previously discoursed on the "success" of the state statute, said:

> There are a few [small tests] that have not come back for one reason or another. It's -- it's not within my knowledge at this point to know if in fact they're given anywhere else anymore. They may just not exist anymore. In any case, the slack has been picked up by other tests.

In other words, the New York sponsors of truth-in-testing legislation have never really bothered to find out about the smaller tests which were displaced by their law. But never mind, the big boys have taken over.

Section 5 (b) of the Weiss bill does exempt tests given to less than 5,000 persons from the provisions of Section 5. However, two things must be noted in this regard. First, Sec. 5 (b) applies to only one provision -- Section 5 -- of the Weiss bill; smaller companies are still covered by the other parts. Second, it should be emphasized that the 5,000 national exemption extended by Weiss is the same as the exemption contained in the amended New York bill for that one state alone. In summary, the 5,000 Weiss exception is of little meaning and effect.

In addition to ignoring the improvements which were made in the New York State law to keep the smaller companies from being unduly penalized, the Weiss bill has added other refinements of its own not contained in the New York statute which are inimical to the smaller test makers. One of these, Sec. 4 (a) (1) (B), requires filing with the Secretary of Education a copy of the contract where one test agency produces or develops a test which is sponsored or administered by another agency. Another, Sec. 7, requires filing with the Secretary the dollar amount of fees received for each test program, the total amount of revenue received from each test program, expenses of the test agency by test program, overhead expenses of the test agency and other proprietary information.

It has never been demonstrated how consumers will be helped by these provisions requiring disclosure of proprietary data. The argument has been made that since public utilities reveal this information, our testing companies should be held to the same standard. I guess that would be fine if testing companies, like public utilities, were under the law also guaranteed a profit.

A Double Standard

It is unfortunate that the proponents of "truth-in-testing" legislation have not seen fit to advocate similar standards for the Federal establishment itself. The Gibbons bill, HR 1312, for example, requires that job entry examinations meet certain prescribed requirements. It seems strange that the Congress should impose these requirements upon the private sector -- requirements which for the most part have not undergone trial -- without first imposing them upon tests given by the Federal government itself.

If the testing practices to be imposed by the pending legislation are, in fact, important to protect "the rights of individuals and the national interests" and "to protect the public interest"* it seems only proper that the very first step should be taken by the Federal government itself in the interests of the many individuals taking civil service examinations each year, as well as the many who take armed service examinations.

In this connection, we would also like to invite the Committee's attention to Sec. 3 of the Privacy Act of 1974 (PL 93-579) which added a new Sec. 522a to Title V of the U.S. Code. Subsection (k) -- "Specific Exemptions" -- of that section permits an agency head to "exempt any system of records within that agency" from certain disclosure and access requirements of the statute. The law includes among these exemptions the following:

> (6) testing or examination materials used solely to determine individual qualifications for appointment or promotion in the Federal service, the disclosure of which would compromise the objectivity or fairness of the testing or examination process;

So what we see here is not only that the Congress is reluctant to apply so-called truth-in-testing standards to the Federal establishment, but it has enacted legislation to exempt Federal agencies specifically from such a requirement. At the very least, one would expect that the proponents of truth-in-testing would ask for repeal of this provision.

---

*See Sec. 3 of HR 1312 and Sec. 2 of 1662

Legislation Does Not Deal With a Principal Problem in Standardized Testing

The "NAACP Report on Minority Testing", published by the NAACP Special Contribution Fund, stated that "it cannot be over-emphasized that the qualifications of the examiner or test user are equal to, if not more important than, the tests which are used, since even the best instrument in the hands of unqualified users can lead to disastrous results." We agree. Unfortunately, this principal issue is not touched by either of the bills pending before this committee. Here is an area where consumers of tests can be helped without imposing Federal controls or running the danger of restricting academic freedoms. It is a field ripe for positive effort.

The AAP Test Committee takes some pride in pointing out that one of the positive recommendations we have made in this area was enacted into law -- Sec. 503(b)(4) of the Education Amendments of 1980 (PL 96-374) authorized teacher centers to provide training to familiarize teachers with developments in testing as well as in curriculum development and educational research. Unfortunately, the effect of this provision could well be diluted by the Education Consolidation and Improvement Act of 1981 (Title V-D of PL 97-35) which consolidates the teacher center program.

What we stress here, however, is that if the Congress wishes to protect the consumers of tests, it must deal with the area which the NAACP report emphasized as "equal to, if not more important than, the tests which are used"; namely, the qualifications of the test examiner or test user. These hearings can never be considered complete until this question is adequately dealt with. Not to do so is to perform a disservice o those who are said to be the objects of the pending legislation's concern.

An Unresolved Issue

After enactment of the New York State law, after which the Weiss bill is generally patterned, the American Association of Medical Colleges challenged the constitutionality of the law in the Federal District Court in Rochester, N.Y. with respect to property rights of the Association in the medical examination.

A Federal district judge issued a temporary injunction against enforcement of the requirements of the law. The AAMC also argued that the New York law requirements violate the Copyright Act with respect to the medical admissions test. Until these issues are resolved, the Committee should refrain from moving on the pending legislation.

HR 1312 (Gibbons Bill)

There follows some comments on specific provisions of the Truth in Testing Act of 1981 (HR 1312) introduced on January 27, 1981 by Representative Gibbons:

1. Sec. 2(3) includes oral examinations within the definition of "test". This could present difficulties both in enforcement and definition. For example, is a person interviewed by a personnel director or a college admissions officer being subjected to a "test"?

2. Sec. 2(4) defines "test score" as a numerical value. Besides the obvious complications with respect to oral examinations, is it the intent of this legislation to mandate that all procedures be reduced to numerical values? This would seem to preclude a simple "pass-fail" or other non-numerical scores. On the face of it, this provision would seem both to be unwarranted given the current statements of need and to have the potential for unanticipated consequences at variance with the central purpose of providing information and insuring equitable treatment.

3. Sec. 2(5) includes in the definition of "persons" the term

"societies". Does this include labor unions? Does this bill govern tests for admittance into labor unions, especially when such admittance controls entrance into an occupation?

4. Sec. 6 raises questions as to who specifically shall be required to furnish the information required in subsections (b) and (c). Admissions procedures to institutions of higher learning and to occupations are represented, in practice, by a variety of arrangements between agencies which produce tests and the institutions, organizations and commerical concerns who use these tests. In the area of occupational admissions, the test producer usually is not directly or indirectly involved in the actual administration of an examination and thus would not be in a position to comply with all the disclosure provisions of this section.

5. Sec. 6 (a)(1) and (2). It is not clear what is meant by a "detailed" description.

6. Sec. 6 (a)(3). The term "reliability" has not been defined. Given the number of interpretations of this concept as it currently applies to tests, it should be more thoroughly examined. Further, the Committee should note that when an examination is prepared for a particular purpose, reliability information may not be available prior to the administration of the examination.

7. Sec. 6 (b)(2). The requirement that individuals be "ranked" in relation to other individuals is inconsistent with Section 6 (c) which prohibits grading of tests of knowledge or achievement on the basis of the score distribution of examinees. Ranking individuals by total test performance may also be contrary to the logic on which the test was constructed and in conflict with currently accepted procedures of considering test scores as elements in the overall profile characterizing an examinee.

8. Sec. 6 (b)(3) requires that the testing agency specify cut-off scores for admission to institutions of higher education or occupations. This is unreasonable both from a technical viewpoint and in terms of a testing agency's capability to fulfill the requirement. Such cut-offs are generally unknown to the testing agency in both academic admissions and occupational testing programs. Decisions on cut-off scores are under the control of the admitting or hiring institution and would logically change with changing levels of achievement or aptitude, size of the examinee population, and available jobs or space. The use of single cut-off scores is discouraged by the testing agencies based on their potential misuse as a mechanistic way of using test information. The use of cut-off scores would also tend to minimize the use of other criteria and would be detrimental to considerations placed on criteria other than test results contained in the totality of a potential entrant's abilities.

When applied in the occupational testing area, the use of cut-off scores would also tend to lead to a situation in which no differential decisions, based on test performance, could be made above the cut-off score. In the long run this situation could eventuate in random hiring of minimally qualified individuals rather than the hiring of those deemed best qualified.

9. Sec. 6 (b)(4) should be deleted because the term "further information" is unclear.

10. Sec. 6 (c). In addition to the conflict between this section and Section 6 (b)(2), this section also calls for the differential treatment of the scores obtained from achievement tests and aptitude tests. Such differential treatment implies a distinction in theory between the way two types of test are treated that is not in agreement with current professional thinking in testing methodology. The requirement that no test which measures knowledge or achievement shall be graded on the basis of the distribution of scores of other test subjects also assumes a static level of ability in our population over time, and/or a static capacity across institutions to accept admissions or hire applicants. Both assumptions are neither desirable nor supportable. Further,

if such a requirement were promulgated in occupational testing, it would assume the existence of a definitive list of minimum skills necessary for job entry. Such definitive lists do not currently exist and their absence would, we believe, lead to pressures at the federal level to mandate what such a list would include; or, in effect, a federal specification of minimum job entry skills. This would clearly be a violation of both institutional and state prerogatives and should be avoided.

11. Sec. 8. The effective date of an act such as this should be no earlier than twelve months after enactment to provide adequate time (a) for the promulgation of the final regulations and (b) subsequent alteration of test materials in order to comply with the law.

In summary, we feel enactment of HR 1312 would result in a myriad of foreseen and unforeseen consequences and would present dilemmas of enforcement and compliance.

HR 1662 (Weiss Bill)

There follows some comments on specific provisions of the Education Testing Act of 1981 (HR 1662) introduced on February 4, 1981 by Representative Weiss and others:

1. Sec. 3 (a)(3) calls for providing the test subject with the correlations between test scores and "success" in a career. Beyond the realization that the presentation of correlations to test subjects is unlikely to communicate information that the test subject can understand, the Committee should note that meaningful correlations between test scores and career success are almost never available due to the difficulty of defining what is meant by "career success". This provision, if implemented for occupational testing, would mandate determining "success" in a career for which admittance is sought. Would this require that each profession or occupation be tested or reviewed over the years throughout the career span in some measurable form so a correlation would be generated between that score and an individual's test score? This provision is unclear and unenforceable in the present context of occupational admissions testing.

2. Sec. 3 (a)(4)(B) requires the comparison of test scores by income group. Information necessary for this analysis if not routinely available would require much more paper work, would potentially be viewed as an invasion of privacy, and would probably generate more misunderstandings about the tests.

3. Sec. 3 (a)(4)(C) requires the expression as a percentage of the improvement in test scores as a result of a test preparation course. The information to form such a statistic is not likely to be available to the test producer sufficient for any meaningful interpretation to be made and would depend almost entirely on the content of a specific test preparation course when compared to a specific test.

4. Sec. 3 (a)(9) refers to "handicapped test subjects", a term which is not defined. The Committee will recall the difficulties that ensued because of the lack of a clear definition of "handicapped" in Sec. 504 of the Vocational Rehabilitation Act of 1973 (PL 93-122). That 43-word section eventually resulted in some 26 pages of regulations.

5. Sec. 4 (a)(1)(B). We can see no reason why any proprietary testing company should be obligated to reveal to its competitors the terms of contracts with its customers. This is an invasion of privacy and one that would place any company complying with this provision at a distinct competitive disadvantage. This provision is not in the New York State statute or in any of the testing bills proposed in the states, and does not belong in any Federal statute. This provision is also violative of the spirit, if not the letter of the Privacy Act of 1974 (PL 93-579).

6. Sec. 5 presents particular problems for proprietary test

publishers. These tests, when used for admittance into an occupation, are frequently administered by individual companies to determine eligibility for employment by the company itself. They are administered frequently, usually many times during a year, and usually to a group of no more than several dozen applicants. As a matter of fact, these tests are quite often sold in quantities that rarely exceed 1,500 copies to an individual purchaser for use over a period of time.

We see two problems, if disclosure of test questions and answers are mandated in the occupational testing area. First, the security of the test, for the purchasing company's purposes, would be breached. It would be unfeasible and uneconomical -- and, considering the quantities, a great bother to the customers -- to substitute a new test to be given within a month or so and to repeat that procedure again and again throughout the year. Beyond financial and administrative considerations, it is also highly probable that no such replacement tests would be available because the cost of test development and data gathering to support reliability and validity would increase to the point where it would not be financially viable to continue the publication of occupational admission tests. Secondly, the tests in question are copyrighted. HR 1662, if applied to occupational entry testing, would force the test publisher to, in effect, give away his copyrighted material substantially before the expiration of its useful life. The entire issue of test copyright should be thoroughly explored by the Committee.

We strongly oppose any applications of Section 5 to proprietary occupational entry tests.

7. Sec. 5 (b) is a mischievous provision. It has been iterated time and again at the hearings that HR 1662 contains an exemption for small tests. However, Section 5 (b) only exempts small tests from the provisions of Section 5 alone, not from the other provisions of the bill. In addition, an exemption for 5,000 nationally, as is pointed out elsewhere in this testimony, is the same as the exemption given for one state alone, New York. Any national exemption should be applied to the entire bill and should be a national-level, not a state-level, figure.

8. Sec. 7 (2). With respect to paragraph (A), (B), (C) and (D), this information is not available to proprietary testing agencies. It should also be noted that these mandatory provisions are not in the New York State law.

9. Sec. 7 (2)(F) would oblige testing agencies to reveal their revenue from each testing program. This is proprietary information which would only serve the purpose of informing competitive companies and would place the reporting company at a competitive disadvantage. This provision is violative of the spirit, if not the letter, of the Privacy Act of 1974 (PL 93-579). Also, this provision is not in the New York statute or in any of the bills proposed in the states and does not belong in any Federal law.

10. Sec. 7 (2)(G) also mandates the release of proprietary information which would only be to the advantage of the competitors of the company providing it and which would place it at a competitive disadvatage. This provision is not in the New York statute or in any of the bills introduced in the states and also does not belong in any Federal law.

11. Sec. 8 (a) is inconsistent with Section 431 of the General Education Provisions Act. Section 431 provides that within sixty days after the enactment of legislation, the Commissioner of Education must submit to Congress a schedule showing when it is planned to promulgate final regulations. These final regulations must be issued within 180 days after the submission of the schedule. It should be noted that the Office of Education has difficulty complying with this provision.

12. If HR 1662 is enacted, the effective date should be no earlier than twelve months after enactment to provide for promulgation of the final regulations and subsequent alteration of test materials to comply with the law.

Since HR 1662 contains major provisions which are not in the New York statute, it cannot be argued that as a result of complying with the New York law testing agencies will have no need for additional time to prepare for compliance with a Federal statute.

In addition, we should like to append to our testimony a brief study "The Weiss Bill and the New York State Law",* which outlines some principal differences between the Weiss bill and the New York law after which it is ostensibly patterned.

Position on Derivative Norms

Publishers of standardized tests develop scales and norms for their tests. These result from the national standardizations that publishers conduct. In contrast, derivative norms are norms for a locally-constructed test that has been equated in some way to the publisher's test so that the publisher's scales and norms can be used for the locally-constructed test.

Because of the close relationship of this issue with some of the discussions at these hearings, we should like to include as part of our testimony the brief AAP "Statement of Position on Derivative Norms". (Attachment B.)

Conclusion

There are those who would wish to abolish all standardized testing. National Education Association Resolution 77-67 states: "The National Education Association strongly encourages the elimination of group standardized intelligence, aptitude, and achievement tests." On the other side of the coin, we are reminded that Irwin Polishook, Vice President of the American Federation of Teachers, asserted that while testing has some flaws, abolishing standardized testing would be "an emotional and damaging strategy" that would "deprive us of a valuable education tool, to which there is no alternative." That is sound advice.

---

*Attachment A

Attachment A

THE WEISS BILL AND THE NEW YORK STATE LAW

According to Rep. Weiss, his bill, the Educational Testing Act of 1981, is modeled after the 1979 New York State statute. In this connection, the following is worthy of note:

1. While the New York statute has been amended twice since its enactment to conform with the experience gained in administering it, the Weiss bill has remained unchanged since it was first introduced in July of 1979.

(a) As originally enacted, the New York law had eight sections. Five of these eight sections have been amended. These sections comprise the substance of the law and deal with definitions, background reports and statistical data, disclosure of test contents, notice, and disclosure of test scores. The three unamended sections each consist of a single two-line sentence and are concerned with regulations, violations and severability.

(b) Two new sections were added to the New York law by amendments, providing for (i) special administrations of tests for religious observers, and (ii) creation of a six-person advisory committee.

2. There are some major differences between the Weiss bill and the New York statute wherein the Weiss bill goes beyond the New York law in its application and severity, as follows:

(a) Sec. 3 (a)(3) of the Weiss bill calls for the test agency to provide the correlation "between test scores and success in the career for which admission is sought". There is no similar provision in the New York law*.

(b) Sec. 3 (a)(4)(B) of the Weiss bill calls for the test agency to compare scores by income groups. No similar provision is in the New York law*.

(c) Sec. 3 (a)(4)(C) of the Weiss bill requires the test agency to report "the extent to which test preparation courses improve test subjects' scores on average, expressed as a percentage". No similar provision is in the New York law*.

(d) Sec. 4 (a)(1)(B) of the Weiss bill requires test agencies to reveal proprietary information, in this case contracts with other test agencies. No similar provision is in the New York law*.

(e) Sec. 5 (b) of the Weiss bill exempts tests administered to less than 5,000 person in a test year. The New York statute is much more complex. Very simply stated, tests administered to less than 5,000 persons must only be disclosed every three years. The Weiss exemption applies to Sec, 5 only,

(f) Sec. 7 of the Weiss bill requires the submission of proprietary information including the revenues and expenses of the test agency for each covered test program. No similar provision is in the New York statute.

(g) Sec. 8 (b) of the Weiss bill provides for a civil penalty of not to exceed $2,000 of each violation. This is four times the $500 penalty provided by the New York law.

---

* No similar provision in the California statute.

Attachment B

STATEMENT OF POSITION ON DERIVATIVE NORMS

This statement has been prepared by the members of the Test Committee of the School Division of the Association of American Publishers (AAP),* which includes most of the major publishers of nationally-normed test instruments.

The Test Committee of the School Division of the Association of American Publishers believes it is important to explain some of the legal and psychometric concerns arising from the preparation and use by others of norms derived from copyrighted nationally standardized tests.

Background of Standardized Testing

It is well known that the development and publication of a standardized test is both complex and expensive. Sophisticated psychometric skills are required to prepare an instrument whose components are so interdependent: item content, response devices, technical analysis and norming.

The integrity, accuracy and usefulness of a standardized test depend upon the quality of expertise used in its construction and the interrelationship of its parts. Functionally, no aspect of the test can be viewed separately from the entirety of its use. A test's norms are as integral to the testing instrument as are its test items. The information which norms convey is only valid to the degree that the test itself is valid and has been administered under proper control.

---

*The Association of American Publishers, Inc. is the major national association of publishers of general books, textbooks and educational materials in the United States.

Standardization is achieved by administering a certain set of test instruments under carefully prescribed and controlled conditions; it is not a "by-product" of testing but at once its means and its end.

Standardization for major achievement and aptitude tests is a huge undertaking, requiring extensive research, detailed documentation, and the cooperation of large numbers of students, teachers, and administrators. Two important functions of a standardization program are to yield a set of norms based on the performance of a sample of school population and to yield sufficient information to judge the adequacy of the selected sample. It is not unusual for hundreds of thousands of children to be tested in order to obtain the samples needed for the norming process. The publisher's investment in a standardization for a major test is a very large part of the total cost of developing the test, the answer instruments and accompanying technical data.

Psychometric Concerns

Recently there has been a trend toward equating tests developed at the local or state level to nationally normed tests. From a psychometric standpoint, publishers have several concerns.

When a publisher prepares norms for its own tests, it represents that a student's standing is based on a large representative sampling of results; the publisher of derivative norms represents that a student obtaining a given raw score on a local test would have achieved a certain national norm standing if that local test had been normed nationally. The use of derived norms with local tests also implies that the local test is the psychometric equivalent of the standardized test, *i.e.*, in overall quality of content and design and in coverage of specific skill areas and levels.

In fact, to properly equate one test to another, the two tests must match each other in both content and levels of difficulty of test item sequence. Furthermore, both tests must have a sufficient and comparable spread of ability over the entire range being tested to ensure valid results.

We do not debate the quality or usefulness of local tests. We *do* believe sufficient evidence exists to justify our view that any assumptions or representations that a local test using derivative norms is the equivalent of the nationally-normed and copyrighted test to which it was equated are unfair to students, the educational community, and finally, to test publishers.

The harm to students when derivative norms are invalid is obvious. Educational decisions about students may be made on the basis of fallacious information. Similarly, the reputation of the test publisher's national norms, and of necessity its testing instruments, can be compromised by the use of invalid derivative norms. Therefore, we believe publishers have both a vested and proper interest in assuring that locally or state-developed tests meet established research design standards before their norms are used for equating purposes.

Legality

Converting local test scores to the national normative data owned by the standardized test publisher is not undesirable but the local or state test developer should remember the normative data are not in the public domain. They are the property of its publisher and the norms themselves are copyrightable under copyright law. Local test programs should not misappropriate the publisher's substantial investment of time, money, and expertise expended in the standardization process.

The unlicensed copying of national testing norms developed by individual publishers is a violation of copyright law. Unlawful copying can occur whenever the data are copied without the publisher's permission, for the purpose of creating a derivative norm, whether the copying is done by hand, by photocopying or by some other method. Also, under the copyright law, inputting normative data into a computer entails copying and therefore violates the publisher's copyright unless explicit permission has been obtained.

Still another kind of copyright infringement occurs when a copyrighted work is revised or recast without the publisher's permission so that a "derivative" work is created.

Therefore, if preparation of a local test involves the unlicensed use of a publisher's national norms, to create derivative norms, then an unlawful "derivative work" results.*

Court decisions protect the rights of publishers in test materials. For instance, a test publisher's copyright in its answer media was upheld in the case of *Harcourt Brace & World* v. *Graphic Controls Corp.*, 329 F.Supp. 517 (S.D.N.Y. 1971). The publication of solutions to test questions was held to be an infringement of the publisher's copyright in the questions in *Addison-Wesley Publishing Co*. v. *Brown*, 223 F. Supp. 219 (E.D.N.Y. 1963), as the court concluded that the product of the publisher's "original and useful effort" had been unfairly used by the defendants.

---

*Derivative Works: Under §101 of the Copyright Act of 1976, the copyright proprietor has certain exclusive rights to its intellectual property, including the right to prepare derivative works:

> A "derivative work" is a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole represent an original work of authorship, is a "derivative work."

Just as the more tests differ, the less reliable is the use of derived norms, the more similar a local test is to a standardized test, the greater is the likelihood of copyright infringement of the test itself. Verbatim copying is not necessary for infringement of copyright of a test. Addison-Wesley Publishing Co. v. Brown, supra.

As a legal matter, the linkage of the name and reputation of the national test to tests of others raises serious questions of trademark infringement and unfair competition under Section 43 of the Lanham Act. See, e.g., Coca-Cola Co. v. Gemini Rising, Inc., 346 F.Supp. 1183 (E.D.N.Y. 1972); Dallas Cowboys Cheerleaders, Inc. v. Pussycat Cinema, Ltd., 467 F.Supp. 366 (S.D.N.Y.), aff'd, 604 F.2d 200 (2nd Cir. 1979).

In summary, we consider that local and state test developers have a professional and legal responsibility to publishers' tests and related normative data. Arrangements for proper legal and psychometric safeguards for both developers and publishers exists in the form of licensing arrangements to ensure proper use and application of publishers' norms. Further information may be obtained by writing to the Association of American Publishers at One Park Avenue, New York, NY 10016 or the individual test publisher.

PREPARED STATEMENT OF TIMOTHY J. MURIS, DIRECTOR, BUREAU OF CONSUMER PROTECTION, FEDERAL TRADE COMMISSION

(The views expressed are those of a member of the staff of the Federal Trade Commission, and do not necessarily represent the views of the Commission or of any individual Commissioner.)

Mr. Chairman and Members of the Subcommittees: I am pleased to provide the views of the Federal Trade Commission staff regarding the possible benefits of coaching for standardized admission tests. My remarks represent the views of the staff of the Bureau of Consumer Protection, and do not necessarily represent those of the Commission or of any individual Commissioner.

Because I assumed my duties as Director of the Bureau of Consumer Protection on October 5, 1981, my involvement with the issues before you began only recently. The Commission's staff, however, has been interested in the effectiveness of coaching for standardized admission tests for some time. In particular, our Boston Regional Office recently concluded an investigation that focused primarily on coaching for the Scholastic Aptitude Test ("SAT"). As you know, that test is used to evaluate applicants for admission to hundreds of colleges and universities across the country each year. The staff initiated its investigation to determine whether commercial coaching schools were making deceptive or unfair claims regarding the effectiveness of their test preparation courses. For example, some coaching schools claimed an ability to raise students' scores on the SAT by as many as 100 points. At the same time, however, the Educational Testing Service ("ETS"), which prepares the SAT, maintained that coaching would do little to improve a student's test performance.

To help determine the accuracy of these conflicting claims, the Commission's staff undertook a statistical evaluation of the effectiveness of two coaching schools that offered programs for the SAT. This research indicated that coaching at one of the two schools raised the SAT scores of its students by an average of approximately twenty-five points each on the mathematical and verbal sections of the test. This finding led the staff of the Boston Regional Office to review other research on the question of coaching for the SAT. The staff found that a growing number of studies have shown that students can significantly improve their SAT scores through coaching.

In light of this evidence, the staff reviewed the descriptive materials about the SAT that ETS, the testmaker, and College Board, the sponsor of the test, provided to students and educators. Although the harsh criticisms of coaching levied in materials distributed in the late sixties had been modified in recent years, the staff found that even the later materials did not appear to recognize the possibility of meaningful score gains through coaching. At the direction of the Commission, these concerns were raised with ETS and College Board. College Board responded by submitting 1980-81 descriptive materials that included a new message to students, providing a more comprehensive explanation of the possible benefits of different forms of coaching.

A more detailed discussion of the research reviewed by the staff, and of the need for improved disclusures to students and educators concerning the benefits of coaching, is contained in the Staff Report entitled "Coaching for Standarized Admission Test." A copy of that Report is appended to this statement. As the Report notes, the emphasis that colleges and universities place on SAT scores lends great significance to disclosure of information regarding the benefits of coaching for standardized tests. Students cannot make informed choices regarding special preparation unless they know that such preparation can affect their scores materially. Similarly, if high schools, colleges, and universities do not know of the benefits of coaching, they may place undue emphasis on standardized admission tests and may err in advising students to forego special preparation for the admission test. In addition, both students and educators could benefit from information about which of the various forms of coaching are most effective for students.

As the foregoing summary indicates, the investigation of standardized admissiontsts focused on assuring that organizations involved inthe standardized testing industry disclose to students and educators the potential benefits of coaching. Of course, the conclusion that coaching can mateerially affect students' scores on standardized tests also raises significant issues of educational policy. For example, the Staff Report notes that the meaning of a test designed to measure skills developed over a lifetime of learning is unclear, if scores can be meaningfully improved in a few weeks or months. The Report also notes that although coaching can materially affect scores, it may not be equally available to all students.

Although the staff's findings highlight these and other issues, they are beyond the statutory mandate and expertise of the Federal Trade Commission. For this reason, the Commission referred these educational policy issues to the Department of Education, the agency with special expertise in this field. These issues also merit serious study by the Subcommittees and by the Congress.

Thank you.

THE NATIONAL PTA,
Washington, D.C., November 13, 1981.

MR. BUDDY BLAKEY,
Subcommittee on Postsecondary Education,
Cannon House Office Building, Washington, D.C.

DEAR MR. BLAKEY: I have enclosed five copies of a statement that the National PTA is very eager to submit for the record on H.R. 1662, "The Educational Testing Act of 1981." Our office has spoken to Mark Pinsky, of Congressman Weiss' staff and it is our understanding that sending you this letter is the proper procedure for submitting such a statement.

If there is a problem, or if you would like us to submit more copies of the statement, please notify me.

Thank you very much for your help.

Sincerely,

SUANNE RUDLEY,
Policy Analyst,
Office of Government Relations.

PREPARED STATEMENT OF THE NATIONAL CONGRESS OF PARENTS AND TEACHERS

Mr. Chairman and members of the subcommittee, we wish to thank you for the opportunity to submit testimony on H.R. 1662, the Educational Testing act of 1981. The National Congress of Parents and Teachers (PTA), an organization of over 6 million members, is deeply concerned about the testing issue.

The PTA endorses the Educational Testing Act as an important step in providing students, parents and educators with more information about standardized tests. The PTA's support for this legislation is not to be construed as opposition to testing, but rather, viewed as a concern that the public be better informed of the nature, purposes, uses, and limitations of standardized testing. The PTA feels that the proposed legislation ensures the proper use of tests by requiring testing companies to disclose information which will help reduce misunderstandings and avoid misuse.

We would like to comment briefly on the major concerns of parents and teachers which have prompted the PTA to endorse this legislation:

(1) Accountability in the testing process

From early childhood through entry into professional schools or jobs, standardized tests make significant judgments about our children. Yet no portion of the public school system is less understood than the facts about standardized tests, their assets and their limitations. The National PTA is deeply concerned that testing programs are being authorized without any broad public discussion of the issues. As parents and teachers, we feel it is important to achieve a greater degree of accountability in the testing process, yet such a review of public discussion is difficult because of the security that surrounds these tests.

Due to the mask of "security," the public has been unable to dispassionately evaluate the claims of test makers. Information about the tests and the testing process is minimal and is rarely shared "after the fact" with those most intimately concerned; students and their families. These tests serve as gatekeepers to the future for our children by developing patterns of expectation by teachers and school systems that often are totally unrelated to a child's true capabilities. Because the power to test is the power to influence curriculum, there is a tendency to reinforce curriculum patterns that respond only to the ultimate goal of "looking good" on the national tests.

An evaluation of assessment requires a great deal more information than is now available to parents and students. The barrier of "security" now keeps us at arms length from the process. That is why we look to legislation like H.R. 1662 as a way of not only accomplishing specific improvements, but as a method of providing us with more information about the testing process and use of the testing results which shape children's lives.

(2) The Purpose of testing: To improve children's education

For years, experts have told us that there are flaws in the present norm referenced standardized testing process and that we must use it because "it is all we have." As parents, we are no longer convinced of that argument, and insist that standardized tests be used to complement the educational process, not dictate a child's educational future. Testing must be responsive to the major goal of a public school system—the improvement of the education of children. We are happy to note that many school systems are moving toward testing systems more in keeping with local goals. Children are being tested in every classroom not so much to compare them with other children in distant school systems, but to determine whether they are learning in accord with the objectives of each local school system, and to determine the effectiveness of the system itself in implementing these goals.

(3) The alternative of national sampling

Sampling on a national basis may be all that is necessary to provide meaningful national comparisons. But nationwide tests are not adequate substitutes for careful local assessments, since the approximately 16,000 local school districts produce a variety of local educational objectives, and different timetables at which specific skills are emphasized. Unless we are planning to move into a national curriculum, these differences of timing and emphasis will continue to exist. A child should not be penalized because a local school system has chosen a different sequence than the test maker. Nor should these important determinations pass from the hands of local school boards who are accountable to their communities to the hands of test producers who are not accountable at all.

(4) A balance between teaching and testing

Is more time being used in a classroom for "testing" than "teaching?" As public school students return to classes they can expect to take more tests than ever before, which leads to questions such as: (a) Is this an effective use of the professional educator's time? (b) Is this the most productive use of a child's limited classroom time? and, (c) Are there ways we could get the answers we need without further disrupting the already fragmented school day?

(5) Testing on what has been taught

The PTA advocates that children be tested on what they have been taught. We need the results of such tests to know whether the system or the child is responsible when learning does not proceed at an expected rate or in a certain pattern. This is essential if there is to be any accountability of local schools and if testing is to be used as a developmental tool so that a teacher can specifically pinpoint the next appropriate steps for each youngster.

(6) Test development

(a) Parents and teachers are concerned about who participates in test construction and the basis on which these people are chosen. This information would be useful in understanding the way in which tests are constructed.

(b) It is critical to know what tests purport to measure so that judgments are confined only to what is measured. Testing companies report that they are blameless for the misuses of test results, yet there are often conflicting points of view among school systems about what test results indicate and what information can be gleaned from the test scores. If it is mandatory for test manufacturers to define the purpose of the tests, parents and teachers will be better able to restrict the use of such data to the purposes for which they were intended.

(c) There is an information void on how standarized tests address the unique testing needs of handicapped and disadvantaged youth.

(7) Test usage

(a) We desperately need information on how tests are used, when and by whom. Discussion is necessary to determine how selective we have been in accepting a whole range of testing to which we subject children.

(b) Multiple choice or short answer types of questions limit the ability to measure a full range of skills. For example, the ability to write well, to organize and transfer thoughts to paper, requires testing of a written sample, yet for many years we have tested writing skills by multiple choice questions. Why? Because it is easier to score. We reduce writing to the rote priciples of grammer, spelling and punctuation, and then wonder why children's writing skills have deteriorated. We must recognize that methods of testing must rely on criteria other than "what the computer can efficiently score."

(8) The responsibility associated with testing

There is no clear indication of who is responsible when national tests backfire or are misused. At present, some problems are caught by random spotting, other are suspected and only verified when a test taker has been able to review the disclosed test questions. Is the test maker responsible? Is the school responsible? Or, is it only the student user?

Certainly, tests are only one measure of a person's ability, but because they are so broadly used, the significant questions about their effect and reliability must be thoroughly aired and other potential measurement techniques carefully considered.

As parents and educators, we support the principles of H.R. 1662 as a means for requiring the disclosure of information which will address the eight concerns just raised and we feel that the following are hallmarks of the legislation:

(a) The bill provides for the standard error of measurement to be shared not only with teachers, but with test takers and their parents. Many parents will, for the first time, be aware of the broad range of precision evident in these scores, even though life and career decisions are being made of the basis of a few points.

(b) The legislation includes protections to the right of privacy of family and test takers in ensuring that research will be limited to data which are not personally identifiable, unless approved by the test taker.

(c) The bill mandates the release to test takers of their actual test questions and answers within a reasonable period following the tests. Also important is a procedure for review of a challenged test score within a reasonable time frame. We are hopeful that the section which will require a description of special services to accommodate handicapped students will allow schools to deal more effectively with the needs of such students for differentiated testing. We are finding, for example, that many students with writing difficulties can be tested appropriately by oral tests. To try to force written tests on such youngsters is cruel and unproductive.

In closing, we would like to emphasize several points:

(1) Charges have been made that these proposed changes will dramatically increase the cost of testing and damage test makers' ability to compile quality test questions. This would be a poor excuse for maintaining a system of testing which is falling under serious challenge. Schools and families now spend over a quarter billion dollars a year on testing, and the appropriate question might well be "are we buying instruments that tell us what we need to know? We cannot evaluate the changes that may be necessary unless we can review what the true costs of the present system cover.

(2) Testing is a large private for-profit industry which has a tremendous impact on a public phenomenon—education. The public has a right to scrutinize this industry and to hold it accountable.

(3) Parents, teachers, and students should be able to review tests and testing materials to ensure that they are understandable. This is particularly important for tests like the SAT and the accompanying disclosure form. There is evidence to indicate that the present method for requesting a copy of the SAT and scored answer sheet is confusing students and is causing a low request rate for this information.

(4) There should be further study as to the effects of coaching on test scores. If coaching has a significant effect, this should be acknowledged, and those who use test scores as an important indicator of ability should be informed.

(5) The National PTA does not seek an end to testing. But we are seriously concerned that the present degree of the art of assessment is far less precise than we have been led to believe. We feel the questions about our present testing methods will not be put to rest until parents and teachers are able to make decisions based on much more information than is now available to us. That is why this legislation is so important. We seek the information that will give all of us dependable data to help determine the course of our public schools and the appropriate direction for our children. The National PTA stands ready to participate in any reasonable effort to reach that goal.

BILL GOODLING
19TH DISTRICT, PENNSYLVANIA

COMMITTEES:

COMMITTEE ON
EDUCATION AND LABOR

SUBCOMMITTEES:

RANKING MINORITY
ELEMENTARY, SECONDARY, AND
VOCATIONAL EDUCATION

HUMAN RESOURCES

COMMITTEE ON
FOREIGN AFFAIRS

SUBCOMMITTEES:

RANKING MINORITY
AFRICA

ASIAN AND PACIFIC AFFAIRS

Congress of the United States
House of Representatives
Washington, D.C. 20515

ROOM 1713
LONGWORTH HOUSE OFFICE BUILDING
TELEPHONE (202) 225-5836

DISTRICT OFFICES:

FEDERAL BUILDING
200 SOUTH GEORGE STREET
YORK, PENNSYLVANIA 17403

CHAMBER BUILDING
212 NORTH HANOVER STREET
CARLISLE, PENNSYLVANIA 17013

POST OFFICE BUILDING
ROOM 209
GETTYSBURG, PENNSYLVANIA 17325

2143 MARKET STREET
CAMP HILL, PENNSYLVANIA 17011

TOLL FREE DISTRICT NUMBER:
800-632-1811

November 16, 1981

Dr. Samuel Messick
Vice President and
  Distinguished Research Scientist
Educational Testing Service
Princeton, New Jersey 08541

Dear Dr. Messick:

Since you are a recognized expert in the area of "coaching" for standarized tests, I have several questions which I would like you to address for the hearing record on the Educational Testing Act of 1981. I would appreciate your prompt reply.

(1) Is the FTC coaching study seriously flawed and, if so, how? Is it possible to use the FTC findings in a technically sound way to estimate the benefits of coaching and, if not, why not?

(2) Is it possible to estimate the benefits of coaching for the SAT from other coaching studies and, if so, what are the likely score improvements attributable to coaching?

(3) Since SAT score improvements associated with coaching seem to increase as student time and effort devoted to coaching increase, would it not follow that students should be coached for as long as necessary to attain the desired score level?

(4) If coaching can improve SAT scores over and above the experiential growth that may occur anyway without the coaching program, are there not important issues of equity in the access to coaching programs?

Your cooperation is appreciated.

Sincerely yours,

William F. Goodling
Member of Congress

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08541

609-921-9000
CABLE-EDUCTESTSVC

Samuel Messick
VICE PRESIDENT FOR RESEARCH

November 30, 1981

DEC 2 1981

The Honorable William F. Goodling
House of Representatives
Longworth House Office Building
Room 1713
Washington, DC 20515

Dear Congressman Goodling:

In connection with the hearings on the Educational Testing Act of 1981, I appreciate the opportunity to answer for the record the four questions addressed to me in your letter of November 16:

(1) Is the FTC coaching study seriously flawed and, if so, how? Is it possible to use the FTC findings in a technically sound way to estimate the benefits of coaching and, if not, why not?

The FTC coaching study is seriously flawed by virtue of employing a research design that cannot control for selection bias or self-selection bias. Selection bias refers to any systematic differences between the coached and uncoached students selected for study that may affect SAT performance -- in addition to the fact that one group was coached and the other was not. If these systematic group differences result from student choice of the coaching program rather than from experimenter choice of the student, they are called self-selection bias. Thus, in the absence of appropriate controls, certain personal factors characteristic of students attending a particular coaching program, such as their motivation or career aspirations, may be responsible, at least in part, for subsequent SAT performance that appears to be the result of the coaching experience.

The standard prescription for avoiding selection bias is an experimental design entailing random assignment of students to coaching treatment groups and noncoaching control groups, for only with random assignment can coaching effects be presumed to be independent of prior status on any of a host of personal or background characteristics. The FTC coaching study did not employ random assignment but, rather, compared students attending two selected coaching schools with uncoached students (selected from College Board test files) who took the SAT during the same time period in the same geographical area. In the absence of randomization, there is an inevitable equivocality in the interpretation of these results because some unmeasured personal characteristics might have influenced both the students' decisions to participate in the coaching program and their subsequent SAT performance, which would tend to inflate the coaching program's apparent effectiveness.

When demographic and personal characteristics of the coached and uncoached groups in the FTC study were contrasted, it was found that the coached group was significantly higher than the uncoached group in high school class rank, parental income, most recent English grades, most recent math grades, and number of years of math taken. In addition, the coached group included significantly more nonpublic school students than the uncoached group. Although appropriate statistical adjustments were made for these and other preexisting group differences for which measures were available, there is no way to take into account unmeasured factors also likely to differentiate the groups, such as student motivation or level of parental education. In the absence of random assignment of students to coached and uncoached groups and especially in view of the large and extensive differences confounded in the nonrandomized groups ultimately analyzed in the FTC study, any score effects derived from these data must be interpreted as combined coaching/self-selection effects. In the absence of random assignment, there is no technically sound way to disentangle the effects of coaching from the effects of self-selection in the FTC study, but some rough estimates of the likely proportion of each might be obtained by contrasting the FTC results with the findings of all other SAT coaching studies, some of which employed random assignment. This approach is discussed in the response to the next question.

(2) Is it possible to estimate the benefits of coaching for the SAT from other coaching studies and, if so, what are the likely score improvements attributable to coaching?

The FTC coaching study is not unique in being seriously flawed -- all of the available studies of coaching for the SAT are flawed in one way or another. Most of the studies were subject to the influence of selection bias discussed in response to question (1), which severely compromises interpretations as to the source or determinants of score effects -- in particular whether they may be unequivocally attributed to coaching experiences as opposed to personal or background characteristics of the (self-) selected students. In this regard, some of the studies involved control groups of uncoached students attending different schools from those of the coached students or else drawn from other extrinsic sources such as test-score files, thereby confounding coaching effects with differential school effects and numerous self-selection factors. In some other studies, control groups of uncoached students were specially constituted to match available samples of commercially coached students on a number of variables, but this still allows systematic differences between the groups on unmatched variables. Another defect common to several studies is an unfortunate reliance on small samples of coached students, which results in imprecise estimates of score effects and a reduced likelihood that real effects will be detected as statistically significant. Moreover, instead of using a regular SAT administration, some studies invoked unrealistic motivational conditions by administering a special SAT that did not count on the college admissions record, which very likely introduced biases in the estimated sizes of score effects. Unfortunately, the three studies using random assignment also employed special SAT administrations which, precisely because they did not count, were probably viewed to some degree as practice tests, thereby eliciting less motivation and effort than would a real SAT administration, especially for the uncoached control students. Finally, and by far the most troublesome from the standpoint of estimating and interpreting coaching effects, some studies utilized no control groups at all. However, although the available studies of coaching for the SAT are methodologically flawed in one way or another, the various defects entail sufficiently divergent implications that any regularities or lawful consistencies cutting across these studies would nonetheless be compelling.

Indeed, a particularly striking regularity was observed cutting across all of these studies -- namely, a remarkably high rank-order correlation, upwards of .7, between the obtained size of score effects associated with coaching and the amount of student contact time involved in the coaching program. That is, the greater the student time and effort devoted to coaching, the larger were the associated score improvements over those obtained by uncoached control students. However, this relationship was distinctly nonlinear, indicating marked diminishing returns in coaching effects, especially for SAT-Verbal. That is, each additional increase in score effects associated with coaching appeared to require geometrically increasing amounts of student time and effort.

By fitting mathematical models to this nonlinear relationship, it is possible to estimate the average coaching effects likely to be associated with coaching programs of different durations. According to these models, an average of 10 SAT score points (on a score scale ranging from 200 to 800 points) would be expected to be associated with about 12 hours of Verbal coaching or about 8 hours of Math coaching; an average of 20 score points would be associated with about 57 hours of Verbal coaching or about 19 hours of Math coaching; an average of 30 points with about 260 hours of Verbal or 45 hours of Math coaching; and, an average of 40 points with about 1185 hours of Verbal or 107 hours of Math coaching. Of course, these figures are not results of any actual coaching program -- rather, they are theoretical estimates of average score gains that might be expected consistent with available data. Nevertheless, these figures may be helpful to our thinking about some of the issues associated with coaching. They suggest, for example, that the time required for motivated students to achieve average score increases much greater than 20 to 30 points over the experiential growth they would have attained without any coaching rapidly approaches that of full-time schooling, especially for Verbal.

(3) Since SAT score improvements associated with coaching seem to increase as student time and effort devoted to coaching increase, would it not follow that students should be coached for as long as necessary to attain the desired score level?

As discussed in response to question (2), the relationship between student time and effort devoted to coaching and the associated coaching effects is distinctly nonlinear, indicating marked diminishing returns in score improvements attributable to coaching. Although in principle it might be possible under these circumstances to coach students for as long as necessary to attain the desired score level, in practice the time required would become unrealistically long. According to the mathematical models fit to the available SAT coaching data, a Verbal coaching program 260 hours long would be associated with an average coaching effect of 30 SAT-V score points over and above what comparable students would have gained during the same testing interval without coaching, while a Verbal program 1185 hours long would be needed to achieve an average score effect of 40 SAT-V points. Thus, for motivated students to attain an additional 30 to 40 SAT-V score points on the average through coaching, somewhere between a third of a school year to more than an entire school year would need to be devoted full-time solely to Verbal coaching. To be sure, greater score effects are observed for SAT-Math, but 107 hours of Math coaching would be associated with 40 SAT-M points and 254 hours with 50 SAT-M points, which still approach a full-time commitment to Math coaching for about a third of the school year. Moreover, the available coaching research suggests that score gains much greater than about 10 or 20 points depend increasingly on genuine skill development as opposed to improved test wiseness. This implies that the verbal and quantitative comprehension and reasoning skills measured by the SAT -- if they are to be systematically developed as abilities facilitative of school and college performance -- might be better developed not by special coaching programs over the short run, but rather over the long run by regular school instruction that integrates the development of thought with the development of knowledge.

(4) If coaching can improve SAT scores over and above the experiential growth that may occur anyway without the coaching program, are there not important issues of equity in the access to coaching programs?

If coaching were to substantially increase test scores without correspondingly improving the skills and abilities measured by the test, there would be important implications for testing practice. Such an outcome would imply that the test or the testing experience entails unintended sources of difficulty that can be at least partially overcome by special preparation — anxiety over being evaluated, for instance, or unfamiliarity with item formats or test-taking strategies. This suggests — even though the available studies find only small average effects, if any, associated with coaching programs stressing test familiarization and practice as opposed to skill development — that in the interest of equity all test candidates should have an opportunity to familiarize themselves with appropriate test formats and to practice recommended test-taking strategies. Such familiarization might be accomplished, for example, through the diligent use of practice tests and advisory materials similar to those routinely distributed by the College Board for the SAT. Issues of equity of access to such test-wiseness coaching programs and materials become important to the extent that student differences in test-taking skills per se influence test scores.

On the other hand, if score increases attributable to coaching represent commensurate improvements in the verbal and mathematical abilities measured by the SAT, then there would be important implications for educational practice and social policy. If coaching techniques for fostering ability development could be specified that were not an integral part of the regular school experience, then the issue of equity of access to special coaching programs arises. It is unfortunate that many of the studies of coaching that have been conducted to date (including the FTC study) were not focussed carefully enough on the content of coaching programs and how they are like and unlike ordinary school instruction. The SAT measures the current level of developed scholastic abilities facilitative of academic learning. Whether this current level of developed abilities derives in part from special coaching programs pointed toward improved test performance or from regular educational programs pointed toward improved school performance or from extensive experiential learning in nonschool settings — that is, from coaching or instruction or experience -- is indistinguishable to the SAT. Thus, the issue of equity of access to coaching programs that are effective by virtue of ability development, if such could be identified, is similar to the issue of equity of access to effective school programs or effective life experiences, for they each have similar consequences for test scores. Such ability-enhancing coaching programs raise important equity issues of educational access, to be sure, but they are not new equity issues nor are they unique to coaching. These are issues that many educators are paying close attention to.

I hope these responses are helpful to your deliberations, and I greatly appreciate the opportunity to contribute them. For the record, I am also submitting three publications in which these issues are discussed in greater detail in the context of available scientific evidence. The publications are entitled: (1) The Effectiveness of Coaching for the SAT; (2) Time and Method in Coaching for the SAT; and, (3) Issues of Effectiveness and Equity in the Coaching Controversy. I am also appending a brief resume of my credentials as an expert in the field of educational and psychological measurement.

Again, thank you for this privilege of participating in the hearing process.

Sincerely yours,

Samuel Messick

Samuel Messick
Vice President for Research

Enclosures

db

Samuel Messick

Samuel Messick is Vice President for Research and Distinguished Research Scientist at Educational Testing Service and Adjunct Professor of Psychology at City University of New York Graduate Center. He earned a Ph.D. degree in psychology at Princeton University in 1954 and has won a number of fellowship awards, including appointment as a resident fellow at the Center for Advanced Study in the Behavioral Sciences at Stanford. He has published over 100 articles in scientific journals and is the author or editor of a dozen books and monographs on such topics as *Problems in Human Assessment* and *Individuality in Learning*. He has been elected President of two major professional organizations in the field of measurement -- the Psychometric Society and the Division of Evaluation and Measurement of the American Psychological Association. He has served on the editorial boards of ten scientific journals, and for seven years he was Editor of the *Review of Educational Research*, a journal of the American Educational Research Association.

THE EFFECTIVENESS OF COACHING FOR THE SAT:

REVIEW AND REANALYSIS OF RESEARCH FROM THE FIFTIES TO THE FTC

by SAMUEL MESSICK

in collaboration with

DONALD L. ALDERMAN, WILLIAM ANGOFF, ANN JUNGEBLUT,
DONALD E. POWERS, DONALD ROCK, DONALD B. RUBIN
EDUCATIONAL TESTING SERVICE • PRINCETON, NJ

AND

T.W.F. STROUD
QUEENS UNIVERSITY • KINGSTON, ONTARIO

Contents

Preface v

I. Overview 1

II. Before the FTC Study:
The Context of Prior Findings on Coaching for the SAT . . . 7

III. Beyond the FTC Study:
Critique of Assumptions and Inferences 31

IV. Estimation of Combined Coaching/Self-Selection Effects in the FTC Study: Detailed Summary and Elaboration of the Stroud Reanalysis 43

V. Disentangling Coaching Effects and Differential Growth: Detailed Summary of the Rock Reanalysis 53

VI. Implications for Educational and Testing Policy and Practice 61

VII. References 69

VIII. Appendices

- **1.** Critical Notes on the FTC Coaching Study 75
- **2.** Stroud Report 97
- **3.** Rock Report 123

Preface

This report examines evidence and arguments about the effectiveness of coaching for the SAT. It views the issue as being much more complicated than the simplistic question of whether coaching works or not. Coaching in and of itself is not automatically to be either rejected or encouraged; it has to be analyzed and evaluated—it matters what materials and practices are involved, at what cost in student time and resources, and with what effect on student skills and attitudes as well as test scores.

The SAT measures developed abilities of verbal and mathematical reasoning and comprehension that are acquired gradually over many years of experience and use in both school and nonschool settings. By virtue of this gradual development, these intellective skills are relatively difficult to improve markedly through brief courses of intervention in the final year or two of high school when the SAT is typically taken. Since these abilities are learned in manifold ways through both instruction and experience, one would expect high quality instruction over extended periods of time to improve them and hence to increase SAT scores. Indeed, score gain across the high school years is the typical pattern exhibited by students taking the SAT. Since coaching at its best is a form of teaching, the key questions are whether the coaching experience is of sufficient quality and sufficient duration to yield significant skill improvement as well as score improvement over and above the experiential growth that would have occurred regardless of the coaching program. If significant improvement requires relatively large amounts of student time devoted to coaching, then the problem becomes expanded to include questions of the difference between coaching and instruction and of the instrumental role of comprehension and reasoning skills in school learning as well as their status as explicit objectives of school learning.

Thus the issue is not just whether coaching works or not, but how much student time devoted to what kinds of coaching experiences yield what level of score improvements in comparison with the level of experiential growth occurring without those coaching experiences. Moreover, since students with different personal and background characteristics often exhibit different performance characteristics and probably even learn in different ways, we should be alert to the possibility that coaching programs, like other forms of teaching, may have differential effects for different kinds of students.

The second printing of this report incorporates the changes previously noted on an errata sheet prepared after the first printing.

Many individuals have contributed to this effort. We would like to thank Thomas Donlon, Garlie Forehand, and Winton Manning for their careful review of the manuscript; Nathaniel Hartshorne for his editorial comments; Rex Jackson and Stephen Ivens for their helpful suggestions on the presentation of arguments and of summary data; Lloyd Bond, Robert Glaser, and Robert Linn for their advice on research and policy implications; and, Anthony Bryk for his comments on one of the applications of the growth model included in the first printing of this report. Special thanks go to John Tukey for his general review of the analysis and its ramifications and, in particular, for his gentle insistence that student contact time in some way or other holds the key to understanding coaching effects.

Samuel Messick
Princeton, New Jersey

I

Overview

This report presents a critique and reanalysis of the Federal Trade Commission's (FTC) study of commercial coaching for the Scholastic Aptitude Test (SAT). The FTC study is one of the largest studies of coaching ever done and one of the few studies of commercial coaching extant, so it merits careful examination. But it is not the only coaching study ever done nor is it free from problems of design, so it should be examined in the context of prior findings.

The first part of the report summarizes the major results of earlier studies in a way that draws special attention to the strengths and limitations of the various study designs. One of the most important of these design features is random assignment of examinees to coaching treatment groups and noncoaching control groups, for only with random assignment can we consider treatment effects to be independent of prior status on any of a host of personal or background characteristics. With random assignment, there are no systematic differences between the experimental and control groups initially and if effective control conditions are maintained, the only systematic difference that will eventuate is that one group will have received coaching and the other will not. In the absence of randomization, as is the case in the FTC study, there is an inevitable equivocality in the interpretation of the results because some unmeasured personal characteristics might have influenced both the student's decision to participate in a coaching program and that program's apparent effectiveness. That is, the same personal factors that led a student to attend coaching school may be responsible, at least in part, for subsequent SAT performance that *appears* to be the result of the coaching program. A number of factors that might lead a person to seek coaching may also by themselves explain why such a person would subsequently perform better than expected on the outcome measure or posttest; for example, a student might not have scored as well on the PSAT or SAT as he or his parents expected in light of high school grades or, in contrast, he might be highly motivated to earn a high score to compensate for a prosaic high school performance. Thus, the effects of self-selection are confounded with effects of the coaching treatment in nonrandomized studies and, consequently, self-selection factors afford plausible rival explanations for the results, or for part of the results, that might otherwise be identified as coaching effects. In such nonrandomized designs, researchers usually attempt to control *statistically* for those potential self-selection factors that have been measured and to analyze the data in alternate ways to assess the sensitivity or robustness of the findings under various plausible assumptions, but there is no way to adjust statistically for self-selection factors that have not been assessed.

An historical appraisal of the effects of coaching on the SAT is complicated because most of the investigations prior to the FTC study were concerned with diverse special preparation programs typically offered by secondary schools. Furthermore, some of these studies used nonrandomized designs and therefore are hampered by the same problems of interpretation that affect the FTC study; some were also poorly controlled and involved small samples. On balance, the average effect associated with participation in a coaching or special preparation program according to those earlier studies that included some type of control group was less than 10 points for the SAT-Verbal score (on a scale running from 200 to 800 points) and less than 15 points or so for the Math score. For example, two studies were conducted using random assignment of students to coaching and control conditions, one dealing with SAT-Math and the other with SAT-Verbal: In a study by Evans and Pike (1973) an average effect of slightly over 16 points for SAT-M was obtained for special preparation involving seven 3-hour sessions with 21 additional hours of homework; Alderman and Powers (1979) estimated an overall special preparation effect across eight secondary schools of about 8 points for SAT-V. However, for particular groups of students and for particular coaching treatments, estimated effects of over 20 points were also reported in various studies. Although substantive content of the coaching programs was not systematically evaluated as part of these studies, the smaller effects appear to be associated with short-term cramming or drill-and-practice and the larger effects, found more often for Math than Verbal, with longer-term intensive programs involving skill development. In addition, results in at least two earlier studies suggested possible interactions for SAT-M as a function of years of math taken and sex. Relatively consistent with this historical context, the FTC study found negligible effects for students attending one commercial coaching school and average score increases of about 20 to 30 points for both SAT-V and -M for students attending another school (where the coaching program entailed 10 four-hour sessions plus homework).

Because random assignment could not be employed in the FTC study, this 20- to 30-point effect for students attending one school is actually an estimate of the joint effects of coaching and self-selection. This confounding of treatment and self-selection effects is one of the major issues addressed in the critique of the FTC study that follows in Section III. It is impossible to determine with confidence whether the effects reported by the FTC study may be attributable in whole or in part to self-selection not controlled under the study design rather than to any impact of the coaching program as such. No statistical reanalysis of the data, given the study design, can fully solve this problem.

Nevertheless, in an effort to contribute further to an understanding of this complex issue of coaching effectiveness, we commissioned two major reanalyses of FTC data. The first one followed an analysis-of-covariance type of model similar to the FTC approach, but included additional background scores as covariates. Furthermore, this first reanalysis introduced an important refinement in that estimated effects were obtained relative to the regression line for noncoached students rather than for pooled coached and noncoached students as in the FTC analysis. This reanalysis covered all three commercial schools in the FTC data set, including one for which no analyses were undertaken by the FTC because of its small sample size. The overall results were similar to the general FTC findings: inconsistent and negligible effects for students at two schools and for students at the third school combined coaching and self-selection effects of about 20 to 35 points for both Verbal and Math scores. In addition, for students attending the two largest schools, this study investigated interactions between the size of effects and the characteristics or backgrounds of the students. Two interactions, one involving race and the other self-reported parental income, were uncovered for students at one of the coaching schools, the one exhibiting negligible effects overall, but not at the other and only for SAT-V, not for SAT-M. The sample of black students was very small and somewhat atypical, but black students at that one school exhibited larger coaching/self-selection effects than nonblacks. Similarly, independent of the race interaction, students reporting low parental income exhibited somewhat larger coaching/self-selection effects than those reporting high parental income.

The second reanalysis employed a statistical model which takes account of differential rates of growth in SAT scores over time, if they occur, for the coached and noncoached groups. However, this analysis does not adjust for self-selection factors or other differences between the groups on background variables unrelated to differential growth. Because this analysis required three test scores for each student, it was undertaken only at the largest coaching school, the one for which the FTC found a correlation between coaching school attendance and SAT performance. Differential growth in the abilities measured by the SAT appeared to operate in these data more for the Verbal score than for the Math score. The resulting estimates of the combined coaching/self-selection effect, taking differential growth into account, were about 17 points for Verbal and 30 points for Math.

After reviewing earlier studies and the FTC study, the present report then considers the implications of the findings for testing practice and policy. One key issue is the extent to which increased test scores attributable to coaching may represent stable long-term improvements in the verbal and mathematical reasoning skills measured by the SAT or instead reflect the overcoming of inadvertent sources of test difficulty unrelated to these reasoning skills, such as difficulty associated with test anxiety and unfamiliarity with different item formats and test-taking strategies. No coaching study, however, has yet directly addressed this basic question of whether obtained score increases reflect stable ability improvements or increased test wiseness.

The SAT measures verbal and mathematical reasoning abilities that develop over many years of experience and use in both nonschool and school settings, and it is difficult to improve them with short-term interventions at seventeen or eighteen years of age, when the SAT is typically taken. Nonetheless, these abilities are learned, and effective learning experiences that facilitate their further development and result in test score increases should be welcomed. Such score increases reflective of improved abilities do not invalidate the SAT—on the contrary, they contribute to the test's construct validity as a measure of verbal and mathematical reasoning abilities and presumably also to its predictive utility, since the improved abilities should serve the student well in criterion situations (such as school or college learning) that entail verbal and quantitative reasoning by analogy, induction, or deduction. Issues of equity of access to coaching programs that may develop these abilities are basically similiar to issues of equity of access to effective school programs or to effective nonschool learning experiences. Thus, any score increases that may represent real improvements in the underlying abilities measured would obviously have implications for instruction and educational practice. No study to date has systematically addressed the question of what kind of instruction may be most effective over what period, for what kinds of students or groups of students, in improving SAT performance. Data that would permit a detailed comparison and evaluation of the commercial coaching programs covered in the FTC report have not been made available.

On the other hand, any score increases that may result from the reduction of construct-irrelevant difficulty on the test have implications for testing practice. Such sources of difficulty should be eliminated to the extent possible through astute test-construction procedures. For this reason also, prospective examinees should be provided with test familiarization materials, sample tests that can be taken at home, and other aids to effective test taking, such as those prepared and distributed by the College Board. Issues of equity of access to coaching programs that help overcome such extraneous test difficulty become important to the extent that student differences in test-taking skills per se substantially influence test scores.

II

Before the FTC Study: The Context of Prior Findings on Coaching for the SAT

The Scholastic Aptitude Test (SAT) was developed as a measure of academic abilities, to be used toward the end of secondary school as a predictor of academic performance in college, and as a standardized supplement to the secondary school record available to college admissions officers. The SAT was explicitly designed to differ from achievement tests in school subjects in the sense that its content is drawn from a wide variety of substantive areas, not tied to a particular course of study, curriculum, or program. Moreover, it taps intellectual processes of comprehension and reasoning that may be influenced by experiences outside as well as inside the classroom. These intellectual skills are exercised to some degree in all subject-matter areas at all levels of schooling, as well as in response to real-life situations. These skills are thereby learned in manifold ways and gradually develop over time as a function of both instruction and experience. The growth of these acquired abilities thus tends to be slower than growth in achievement areas such as French, chemistry, ancient history, and trigonometry, which are primarily learned as a function of curriculum and instruction. In addition, the SAT is intended as a prospective measure for the prediction of future academic performance rather than as a retrospective measure for the evaluation of past school-related attainment. The specific item content on the SAT attempts to sample the sort of cognitive skills underlying college-level performance: reading with comprehension, understanding vocabulary, verbal reasoning, computational skills, quantitative reasoning, and problem solving.

Thus, the purpose of the SAT is to predict academic success in college. It was designed to reflect verbal and mathematical abilities that are acquired over an extended period and hence can be expected to be difficult to improve substantially through short-term instructional efforts. However, these reasoning abilities are learned and if effective learning experiences, including effective coaching programs, facilitate their development, one would expect not only increases in SAT scores but also increases in criterion performance—in this case, performance in college. Such increases in test scores and academic performance do not invalidate the SAT—rather, they contribute to its construct validity and predictive utility. Coaching or special preparation programs that increase test scores that were inaccurately low because of anxiety, for example, or test unfamiliarity would also improve the validity of measurement by virtue of reducing the prior invalidity of measurement. If, on the other hand, special preparation or coaching efforts result in a substantial increase in test scores without a corresponding positive effect on the level of verbal and mathematical reasoning abilities measured by the test, the value of the test scores for admissions purposes would be undermined.

Thus, three major potential effects of coaching may be distinguished: First, some coaching programs may improve the abilities and skills measured by the test, resulting in commensurate increases in test scores; second, some coaching programs may enhance test-taking sophistication or reduce anxiety associated with taking tests, resulting in increased test scores that are more accurate assessments of ability and skill; and, third, some coaching programs may teach test-taking strategies and answer-selection tricks, resulting in increased test scores that are inaccurately high. Some coaching programs, of course, may produce none of these effects or more than one in various combinations. The first two effects, if they were realized, would be good from the standpoint of the student and good from the standpoint of test validity. The third effect is probably minor with well-constructed tests, because test-makers should strive to minimize the use of complicated or tricky item formats and to eliminate items that may be answered on the basis of clues unrelated to the abilities tested. Moreover, test-makers should also strive to reduce the import of the second possible effect of coaching by providing test-familiarization materials and practice tests to all candidates as well as advice on guessing, reviewing, pacing, and the like, in part to reduce apprehension about what to expect. With respect to the first type of effect, if coaching or special preparation programs can improve the abilities measured by the SAT, the effective components and techniques should be identified and incorporated more widely in secondary- and possibly elementary-school instruction. There is little question that the verbal and mathematical abilities measured by the SAT are learned, but there are large questions about how they can be taught.

There is a progression or ordering of educational tests ranging from measures of scholastic and intellectual abilities at one pole, with content drawn from a variety of substantive areas, to measures of academic attainment at the other pole, with content specialized by subject-matter field. The SAT falls toward the first extreme—it taps general intellectual processes that develop gradually over many years of experience and use both inside the classroom and in everyday life, and these processes should therefore be relatively difficult to enhance markedly through brief courses of intervention. The typical educational achievement test falls toward the other extreme—it taps specific knowledge and skills acquired through the normal course of classroom instruction or independent study, and this knowledge and skill should therefore be relatively responsive to instructional intervention, even in brief courses. Similarly, there is a progression of types of preparation for taking examinations ranging from simple practice on sample items at one extreme to the provision of intensive instruction aimed at developing ability and knowledge at the other extreme. What has come to be called "coaching" is now generally considered to fall anywhere in the broad range between these two extremes, entailing some combination of test familiarization, drill-and-practice with feedback, training in strategies for general test-taking and for specific item formats, subject-matter review, and skill-development exercises—although widespread usage of the term in the past tended to underscore drill and cramming toward the practice end of this range. The important point, however, whatever the special test preparation is called, is whether or not it leads to significant test score increases and, if it does, whether those increases represent genuine long-term improvements in the knowledge and abilities measured by the test as opposed to enhanced test-taking skills irrelevant to the purposes of testing. The studies that have been undertaken in this area, including the FTC study, have tended to focus on the prior question of whether or not special preparation may produce score increases. Because of the relatively modest effects identified, there has been little impetus to consider whether any obtained score increases represent improved abilities that endure over a long term.

Although there was considerable variation in the score increases observed for particular groups of students and for particular programs or treatments, the earlier studies on the average reported score increases associated with special preparation, relative to score increases for control groups, of less than 10 points on SAT-Verbal and less than 15 points on SAT-Math, on a score scale ranging from 200 to 800 points. Studies lacking control groups yielded larger effects but since they differed from control-group studies not only in design characteristics but in critical program characteristics, their interpretation is especially problematic. In spite of an apparent consistency of results when some type of control-group comparison was included, many of these studies had methodological weaknesses that detract from the strength of the conclusions drawn. We now summarize the results of these earlier studies and consider in detail their particular strengths and weaknesses. The FTC study will be analyzed in the next section of this report.

Studies With No Control Groups

Three studies have been conducted—those by Pallone in 1961, by Marron in 1965, and by Coffman and Parry in 1967—that lacked any control group for evaluating unusual patterns of score change. Pallone (1961) looked at the effects of short- and long-term intensive developmental reading courses on SAT scores of students in a private school for boys. The courses were undertaken "for students in their final year of pre-college work, including a large number of high school graduates who were completing a year of post-high school study in preparation for entrance into the U.S. government academies" (p. 655). According to Pallone, to improve the skills measured by the SAT, "not 'coaching' methods, but instruction of a developmental nature in reading and vocabulary skills was indicated. Improvement in scores could be expected only if the basic skills measured by the test were first strengthened" (pp. 654-655). This program provided focussed instruction to strengthen reading achievement along with intensive practice in reading skills, including such special skills as skimming and critical reading, as well as a brief analysis of typical verbal analogy test items. Approximately 20 students participated in a six-week summer pilot program that met for 90 minutes daily. Substantial percentile increases corresponding to mean reading achievement were reported on the basis of pre- and posttest results. An average score increase of 98 points was obtained on SAT-Verbal from the March 1959 to August 1959 administrations. The long-term program covered a six-month period from September to March with daily meetings of 50 minutes each. From March of 1959 to March of 1960 for some 80 students who completed the long-term course, an average SAT-V increase of 109 points was reported, although the difference in mean scores in Pallone's Table 3 (p. 656) was only 84. The students who participated in the summer program also completed the long-term reading course. The mean increase in their Verbal scores for the period from March 1959 to March 1960 was almost 122 points, or an average of about 24 points over the increase reported after the summer course. The special quality of the sample and the lack of control groups severely limit the implications of these findings vis-à-vis coaching. Furthermore, the instructional focus on skill development and the intensive and long-term nature of the programs put Pallone's efforts close to what ordinarily would be considered "instruction" in contradistinction to "coaching," as Pallone himself insisted.

In the absence of control groups of similar students at this preparatory school who were not taking Pallone's course, it is difficult to assess the import of these score gains. Pallone (1961) suggested comparing them with normal expectations of gains of about 35 points on SAT-V during the final secondary school year, which would yield an instructional or program effect of about 83 points for the summer pilot program (prorated for the five months between tests) and about 74 points for the long-term program. This is not a very satisfactory comparison, however, since Pallone's students were not representative of students in their final year of high school who take the SAT. Slack and Porter (1980) suggested comparing Pallone's results with average gains in national administrations of junior- to senior-year retesters having the same initial average score levels as Pallone's students, which yields prorated instructional effects of 85 and 79 points, respectively, for the summer and long-term programs. Again, this is not a very satisfactory comparison because Pallone's private school students were not a representative sample of the national population of test repeaters. Pike (1978) suggested comparing Pallone's results with average gains of control students in superior schools from other studies of proprietary programs—for example, those conducted by Frankel (1960) and by Whitla (1962) described below. When adjusted for time differences between pre- and posttests in these studies, this comparison leads to instructional effects of about 75 and 53 points, respectively, for the summer and long-term programs. If compared with score gains of control students in other coaching studies who had average initial score levels roughly comparable to Pallone's groups, namely in this instance a selection of control schools from the study by Dear (1958) described below, the adjusted instructional effects are 80 and 65 points, respectively. The point is that in the absence of comparable control groups, no generally satisfactory estimate of instructional effects can be obtained. If an average is taken across the four adjustments just suggested, the resulting estimates of instructional effects are 81 and 68 points, respectively, for the summer and long-term programs. These values are very likely still overestimates, however, because none of the comparisons can take into account the highly self-selected nature of Pallone's private school students, many of whom were completing a post-high school year highly motivated to increase their chances of entering service academies or selective colleges. But given the overall size of the effects, even with somewhat larger adjustments, it seems likely that Pallone's (1961) intensive summer and long-term efforts at "instruction of a developmental nature" may have succeeded to some degree in strengthening basic skills measured by the SAT.

Marron (1965) examined SAT score gains for students at ten well-known preparatory schools that specialize in preparing high school graduates for admission to the service academies and selective colleges. The instructional programs entailed "six months of full-time exposure to course content that is directly related to the verbal and mathematics College Board tests (both Aptitude and Achievement)" (p. 1). A special administration of SAT-V and -M and College Board achievement tests (English composition and intermediate or advanced math) at all ten schools in September 1962 served as the pretest, while the posttest was the regular College Board admissions testing administration in March 1963. It should be noted that if the level of motivation and effort on a special pretest that did not count for college admission was not comparable to that on the regular posttest, the instructional effects in Marron's study would likely be overestimated. Since significant differences were obtained among the ten schools with respect to both the September pretest scores and the March posttest scores and these latter differences remained significant in analyses of covariance adjusting for pretest levels, the overall results were reported separately for groups of schools having nonsignificant differences within group. Score gains on SAT-V were 77 points for group 1 (2 schools, N = 83), 56 points for group 2 (6 schools, N = 600), 47 points for group 3 (1 school, N = 5), and 35 points for group 4 (1 school, N = 26); the weighted average SAT-V increase over all groups was 58 points. Score increases for SAT-M were 83 points for group 1 (4 schools, N = 232), 78 points for group 2 (3 schools, N = 405), and 72 points for group 3 (3 schools, N = 78); the weighted average SAT-M increase over all groups was 79 points. Weighted average gains on achievement tests were 83 points for English composition, 133 points for intermediate math, and 127 points for advanced math. Again, in the absence of control groups, it is difficult to appraise the size of these instructional effects, but it appears that on a relative basis achievement test scores improved much more than SAT scores—on the average about half again as much.

Marron (1965) suggested comparing the SAT score gains with those considered typical for males in their senior year in secondary school, which he reported based on College Board data as 40 points for SAT-V and 43 points for SAT-M over a 10-month testing interval. Prorated for the six-month testing interval in Marron's study, this yields an adjusted weighted average of 34 points for SAT-V and 53 points for SAT-M. As was done for Pallone's (1961) results, the suggested adjustments of Slack and Porter (1980) and Pike (1978) were also applied to Marron's figures, along with an adjustment based on score gains of control students in other coaching studies who had average initial score levels roughly comparable to Marron's groups. If an average of all four of these adjustments is taken, the resulting weighted average values are 35 points for SAT-V and 54 points for SAT-M. But again, none of these adjustments is very satisfactory because none of the suggested comparisons takes into account the highly self-selected nature of Marron's students, thereby leaving important factors of differential motivation and growth uncontrolled. In any event, the relevance of Marron's (1965) study to the issue of "coaching" is arguable, since six months of full-time exposure to course content directed at verbal and mathematical knowledge and skills would ordinarily be considered "instruction." These score gains, whatever comparison groups they are contrasted with, might better be interpreted as testimony to the learning and skill development that highly motivated high school graduates can accomplish with six months of full-time concentrated effort on focussed curricula.

In an attempt to explore further the effects of developmental reading instruction on SAT-V scores, especially in light of Pallone's (1961) findings, Coffman and Parry (1967) undertook a study of three groups of college freshmen who took the SAT-V before and after completing a course in accelerated reading. The course was described as stressing speed of reading with relative accuracy. Pre- and posttest scores based on special administrations of the SAT were available [illegible] two small groups of 10 and 9 students who elected an eight-[illegible]eek course meeting six hours each week. Pre- and posttest SAT-V scores were also available for

25 students whose course met three hours a week for 15 weeks. For the eight-week course, the increases in SAT-V were 3.5 and 9.9 points, while a 28.9 mean loss was observed for the group taking the 15-week course. This score decrease in the 15-week course may stem from problems of test administration and score equating since time constraints dictated the use of a shortened SAT-V for that group. It may also reflect problems in the motivation of students who take a special SAT when they are already in college, as may the relatively modest score increases in the other two groups—although all of these students were presumably motivated to enroll in the course, which explicitly entailed taking the SAT. In addition, the lack of control groups is again a factor that seriously impairs the usefulness of these results. Although these findings are in sharp contrast to the score increases reported by Pallone, the accelerated reading program studied by Coffman and Parry appears to be considerably different from Pallone's developmental reading curriculum and apparently not as directly relevant to the skills measured by the SAT.

Studies With Nonequivalent Control Groups

A second methodological weakness obtains in three other earlier studies: Although each incorporated control groups, these groups were drawn from schools different from those providing special preparation for the experimental students, thus confounding treatment or preparation effects with school effects.

Dyer (1953) studied seniors at two highly selective independent schools for boys—225 students at one school served as the treatment group and 193 students at the other school served as the control group. The students at both schools took the SAT twice, once in September 1951 and again in March 1952. The two groups of students were similar with respect to length of time enrolled in the school, level of SAT scores, and the level and number of foreign-language and math courses taken. The experimental group completed 12 verbal practice exercises in 30- to 60-minute sessions and five math practice exercises in 60- to 90-minute sessions. The control variables in an analysis of covariance were initial SAT scores, number of years each student had been enrolled in school, and the number and level of foreign-language and mathematics courses taken in the senior year. The estimated increase in score for the treatment group over the control group was about 4½ (4.6) points in Verbal and about 13 (12.9) points in Math.

When the students were divided into those who were not taking mathematics courses as seniors and those who were, the no-math boys who were coached gained over 29 points more than those who were not coached. In contrast, it was found that the boys taking mathematics who were coached gained 3.3 points more than those taking mathematics who were not coached. Dyer's conclusions were (1) that coaching on SAT-Verbal is not likely to be effective; (2) coaching on SAT-Math will be of some advantage *only* if the students coached are not already enrolled in regular math courses.

In 1955, John W. French reported a study conducted at three schools: The students at School A (N = 158) pursued their regular courses with no attempt at special preparation for the SAT and served as a control group for coaching in both Verbal and Math; the students at School B (N = 110) served as a treatment group for Verbal and as a control group for Math; and, the students at School C (N = 161) served as a treatment group for both Verbal and Math. The special preparation program in Verbal at School B differed from the one at School C in that the former primarily emphasized vocabulary for a total of 4½ hours, while the latter reviewed 10 verbal exercises more representative of SAT-V skills. The students were all seniors who planned to go to college. In addition, some students in School C participated who were enrolled in a technical course. All students took the SAT in September 1953 and again in March 1954. The data were subjected to an analysis of covariance, using SAT scores earned in September 1953 as the independent or control variable and SAT scores earned in March 1954 as the dependent or outcome variable. The results of French's analysis were mixed. The advantage in Verbal score for boys and girls combined who *were* coached over those who were *not* coached was found to be 18 points in one school and 5 points in another (the one with the vocabulary coaching program). The score increase in Math for boys and girls combined who were coached over those who were not coached was shown to be 6 points when compared to one control school and 18 points when compared to the other. Thus, the largest benefit resulting from coaching for groups of boys and girls combined was 18 points in Verbal and 18 points in Math. When broken down by sex and current enrollment in math courses, the SAT-M data indicated that coached boys not studying math at the time of coaching showed greater increases over control students than did coached boys who were studying math, by about 4 points in one school and 10 in the other. This pattern of higher coaching effects for boys not currently studying math is consistent with the Dyer results. In contrast, coached girls not studying math at the time showed *smaller* increases over control students than did coached girls who were studying math. Coached girls currently studying math exhibited score increases of about 20 and 30 points over the non-coached girls currently studying math in the two control schools, whereas the coached girls not studying math exceeded their non-coached counterparts in the two math control schools by only 1 and 4 points.

French also investigated the effects of using some identical items in coaching and on the posttest. Shortly before the March SAT administration, the students in School C were given a practice test containing 50 verbal and reading items and 34 math items, half of which had been included in the previous practice exercises of the coaching course. Students at a fourth school who received no special preparation served as a control group. The raw scores on the familiar and unfamiliar halves of the test were each converted to the SAT 200- to 800-point scale, and in an analysis of covariance the score on the coached items was predicted from the score on the uncoached items. A comparison of the results showed an increase of 15 points on SAT-M and 47 points on SAT-V for the coached over the uncoached students on the projected full-length test made up of items explicitly coached. These findings relate to a test made up of items identical to those on which students received coaching, and it is unlikely that this situation would ever arise in practice. Nevertheless, in 1974 June Stern included an identical subset of items in the April SAT administration and again in the November administration with no experimental coaching intervention. She found that the average net gain in Verbal as well as in the Math score was approximately 17 points for repeaters of identical items over gains for repeaters of non-identical items. Clearly, if just taking an identical subset of items twice yields a 17-point increase, the effects of *coaching* on identical items, if one had access to them, could be substantial.

In 1958, Robert Dear reported a study to determine whether longer periods of coaching in small groups—two class periods a week for 6 weeks and for 12 weeks—were likely to be more effective than the shorter-term coaching studied by Dyer and by French. In Dear's study, 6 public and 4 private secondary schools were chosen randomly from a list of schools from which at least 15 students had taken the SAT as juniors in May of 1956. A treatment group from each school was selected at random from students who volunteered for coaching. A second group of nine schools from the same geographical region—the New York-New Jersey-Greater Philadelphia area—was drawn at random as the group of control schools. All students eligible for the study had indicated an interest in being coached. Three students were selected from each school from each of three ability levels—90 coached students and 81 control students. Of these, 71 coached and 79 uncoached students took the SAT in May 1956 and in March 1957. The coaching program began in mid-November and continued through mid-March with weekly coaching sessions supplemented by one additional hour of homework each week. Most students repeated the SAT in January 1957, halfway through the coaching period, and again in March 1957. The January results showed about a 22-point advantage for 60 coached students on Math, but a 2½-point disadvantage on Verbal. The control groups for this January administration included all uncoached students (N = 526) in both the same schools and the control schools who indicated interest in being coached. The March advantage for the coached students on Math was about 24 points. The Verbal score results, unfortunately, were not determined because of a significant difference in the slopes of the regression lines for the coached and uncoached groups.

Studies With Matched Control Groups

A third methodological problem occurs in studies in which control and experimental students, although from the same school, are not assigned randomly but are matched on selected measures, thereby still permitting systematic differences between the groups on other nonassessed variables. In 1960, Edward Frankel published the first study on the effects of commercial coaching on SAT scores. Frankel selected 45 high school students who had taken commercial coaching courses and matched them with 45 control students from the same high school in the following ways: (1) their pretest scores on the May 1958 SAT were within 10 points of each other on both Verbal and Math; (2) they had taken the same form of the SAT in either December 1958 or January 1959 to serve as a posttest; and, (3) they were of the same sex. Within each pair, one had taken a commercial coaching course involving roughly 30 hours of coaching in classes of about 25 students between May 1958 and January 1959. In brief, Frankel found an 8.4 point advantage for coached students on Verbal and a 9.4 point advantage on Math.

In 1962, Dean Whitla compared the score increases of 52 students who had attended an intensive 10-hour course in improved study habits, reading skills, and math concepts at a proprietary school in Boston with the score increases of a comparable group of 52 students from the same area who had not taken a coaching course. All of the students had taken the SAT in March or May of their junior year and, in addition, were administered a special SAT in the fall of 1959 when the study began; the January 1960 SAT served as the posttest. The average V and M scores of the two groups were within one point of each other on the spring SAT and within two points on the fall pretest, suggesting that the two groups were not only well-matched in terms of initial level but also in terms of growth rate over this period. Whitla found an 11-point advantage for the coached group on Verbal but a 5-point disadvantage on Math when the posttest was compared with the fall pretest; there was a 10-point advantage on Verbal and a 7-point disadvantage on Math when compared with the spring SAT.

Studies With Randomized Control Groups

Three studies employing a randomized design have been carried out. The first one, by Roberts and Oppenheim (1966), utilized the Preliminary Scholastic Aptitude Test (PSAT) as both pretest and posttest. In contrast to earlier coaching studies that involved highly selective and effective private schools and specialized or suburban public schools, the Roberts and Oppenheim study was undertaken to investigate whether students receiving less adequate instruction might especially benefit from special preparation. Data were collected from 18 predominately, if not entirely, black secondary schools in rural and urban Tennessee. In 6 schools coaching consisted of special instruction in verbal material, in 8 schools coaching was for mathematics, and in 4 schools no special instruction was provided. Within the treatment schools students were assigned randomly to coached and uncoached groups. The instruction was provided in 15 half-hour sessions over a 4- to 6-week period. The results showed small increases for the coached groups over the control groups: about 1½ (1.44) points on PSAT-Verbal and less than 1 (.81) point on PSAT-Math, increases which correspond to about 14 points on SAT-Verbal and 8 points on SAT-Math. This advantage of coached over control groups, however, was due as much to score decreases on the part of the control students, possibly signaling problems in motivation or attrition, as to score increases on the part of the coached students.

A second randomized study, conducted by Evans and Pike (1973; Pike & Evans, 1972), examined intensive coaching efforts in the math area. A sample of 509 students in 12 schools participated in the study. The coached students received 21 hours of instruction and 21 hours of homework, over a 7-week period during November and December 1970, directed at one of the following item types: Regular Math (RM), Data Sufficiency (DS), or Quantitative Comparisons (QC). Three randomly chosen groups of students were defined in each school: one to be instructed in QC, one in either RM or DS, and one as a control group. These groups took the SAT first in October 1970, which served as a pretest, and again in December 1970 (posttest) and April 1971 (delayed posttest). The pretest and posttest were special administrations of the SAT, whereas the delayed posttest was a regular administration. The three experimental groups were given special preparation during November and December, and the control group received instruction after the December posttest. On this schedule, all groups received instruction in test-taking skills—becoming familiar with test directions, pacing, and appropriate strategies for using partial information and guessing. All groups also had instruction in math content—numerical facts, numerical and basic algebraic skills, and in particular mathematical areas such as inequalities. In addition, there was practice on one of the item types for students in each respective experimental group. The study revealed score increases beyond those experienced by the control group for each of the three experimental groups coached on a particular item type. However, because the Evans-Pike study was designed to investigate the relative susceptibility of three item formats to special instruction, it is difficult to say just how large the effects were in terms of SAT scores. The authors' best estimate of score increases reflective of coaching for all four groups over the total period from October to April was about 25 points. The average increase over the control group for the three experimental item-type groups, weighted according to their respective sample sizes, was 16.5 points over the period from the October pretest to the December posttest. Pike (1978) later conjectured that, still keeping within the total 21 hours of special preparation, a judicious combination of instruction for both RM and DS, the two major item-types then in SAT-M, would be expected to yield coaching or special preparation effects of about 33 points.

In the third randomized study, Alderman and Powers (1979) investigated the effectiveness of existing secondary school programs that had been initiated by the schools to improve the performance of students on the SAT-Verbal scale. Students at each of eight schools for whom PSAT scores were available were randomly assigned to a special-preparation group or to a control group. Access to the same preparation course was delayed for control students for the purpose of this study. A special administration of a retired SAT was used as the posttest. Across the eight schools the overall increase in SAT-V attributable to special preparation was about eight points, which is statistically significant at the .05 level. The actual effects ranged from -3 points at one school to +28 at another. Differences in effectiveness between the coaching programs were not statistically significant from school to school, however, and the best statistical estimate of the range was from 4 to 16 points. Nor, apparently, did the control groups react in comparable fashion from school to school, possibly reflecting differences in motivation or seriousness in approaching a special SAT. The largest school effect of 28 points, for example, resulted in part from a control-group decrease almost equal in magnitude to the treatment-group increase (-11 score points in going from the converted PSAT to the SAT for the control group versus +13 score points for the treatment group, which yields a 28-point school effect after covariance adjustments are made). An attenuated form of this pattern occurred at two other schools, whereas the remaining five schools showed varying degrees of score increase for both treatment and control groups. Overall, only one-tenth of one percent of the total variance in SAT-V scores was accounted for by special preparation alone, while school attended taken alone accounted for 24 percent of the Verbal score variance. Furthermore, scores on the PSAT-V and the Test of Standard Written English accounted for 54 and 39 percent of SAT-Verbal score variance, respectively. The authors concluded that the school attended and scores on earlier tests of verbal and writing ability contribute much more to determining SAT-V scores than did special preparation as typically offered by secondary schools.

Another, more subtle methodological problem has emerged in the process of reviewing these coaching studies having randomized control groups, namely the problem of engendering and maintaining realistic motivation for taking the posttest SAT, especially for uncoached control students. Developing realistic motivation and effort for taking pre- and posttests is a common requirement of all coaching studies. However, the three experiments that employed randomized control groups happened also to use as the posttest a special administration of the SAT or PSAT rather than a regular administration. These special administrations may have been viewed to some degree as practice tests that did not count for the record, thereby eliciting less motivation and effort than would a regular SAT administration. Warning signals suggesting this possibility were noted in the Alderman and Powers (1979) study, in which control groups in three schools were found to exhibit score *decreases* in going from a regular PSAT to a special SAT, even though the expectation is for a score *increase* from an October PSAT to a Spring SAT of upwards of 10 or 12 points, which is the basal estimate provided by national administration samples. Such control-group score decreases may substantially complicate the interpretation of the results. One way of viewing the Alderman and Powers analysis, for example, is that their appraisal of the import of within-school intercept differences tacitly implies that within each of the eight schools the effect on student performance of taking a special rather than a regular SAT operates consistently for both uncoached and coached students on the average. Differences in performance between these groups at a given school may then be used to estimate coaching effects on the implicit assumption that any reduced motivation and effort on the special SAT posttest would have lowered mean performance levels of both groups to about the same degree. Such assumptions are not as apparent when the analysis is viewed in terms of sampling fluctuations across eight independent samples: The control-group score changes from pretest to posttest ranged from -11 to +30 with a mean of 8, and the treatment-group score changes ranged from 3 to 44 with a mean of 17; the 9 point overall advantage for the coached over the control students, when adjustments for covariates are made, corresponds to the 8 point coaching effect reported in the study.

Control-group scores also decreased in the Roberts and Oppenheim (1966) study, which similarly employed a special administration posttest, in that case a special PSAT. Unlike Alderman and Powers, who used a regular administration of the PSAT as a pretest, the Roberts and Oppenheim pretest PSAT was a special administration as well. None of the nonrandomized coaching studies reviewed earlier displayed control-group score decreases, and the posttest in all of those studies was a regular administration of the SAT—although Dyer (1953), French (1955), and Whitla (1962) did use special pretest administrations and thereby introduced the possibility of other biases. The remaining randomized study, conducted by Evans and Pike (1973), employed special administration SATS as both pretest and posttest, but their delayed posttest was a regular SAT administration. Pike's (1978) subsequent interpretation of score increases from posttest to delayed posttest as reflecting the long-term consolidation or continuance of gains due to coaching becomes jeopardized from this vantage point by the plausible rival interpretation that those score gains instead reflect increases in motivation and effort in going from a special administration to a regular one.

Comparison of Results Across Studies of Coaching

For numerous reasons, including the diversity of design limitations and the differences in sample sizes, it is difficult to compare results across these several studies in a meaningful way (cf. Pike, 1978). Table II-1 represents one such attempt for those studies having some type of control group. The size of coaching effects reported there were calculated uniformly as follows: When analysis of covariance was performed, the reported values are intercept differences between the experimental and control regression lines, weighted in the case of multiple experimental or control groups by their respective sample sizes. In four studies not reporting analyses of covariance, the values in Table II-1 are average score increases of experimental over control groups, again weighted in the case of multiple experimental or control groups by their respective sample sizes. Two of these latter studies (by Frankel and by Whitla) involved statistical matching, and two (by Roberts and Oppenheim and by Evans and Pike) involved randomization. Averaging these results over all of the studies in Table II-1, weighting in each case by the size of the experimental sample, yields 9.1 points for Verbal and 13.0 points for Math (the unweighted averages are 8.5 points for Verbal and 12.3 points for Math).

For those studies having no control groups, a summary is provided in Table II-2. The special preparation programs summarized there focus on verbal or mathematical content knowledge and skill development and entail the largest amounts of student contact time of any of the studies reviewed, which would ordinarily lead one to characterize them as instruction rather than coaching. These programs range from 45 hours of student contact time over 6 weeks to 48 hours over 8 weeks to roughly 100 hours over 6 months to virtually full-time over 6 months or approximately 600 hours. In contrast, the most intensive of the control-group stud-

Table II-1

Average Difference Between Experimental and Control Groups in Studies of SAT Interventions (Adapted from Donlon)

Study/Design	*Sample Characteristics*									
	School	*Level*	*Sex*	*Characteristics of the Special Preparation*	*SAT-Verbal Difference*[1]	*Significance Level*[2]	*N Exp./Control*	*SAT-Math Difference*[1]	*Significance Level*[2]	*N Exp./Control*
Dyer (1953) Control, different school	Private	H.S. Seniors	M	Twelve 30-60 minute sessions for verbal, five 60-90 minute sessions for math	4.6	.05	225/193	12.9	.01	225/193
French (1955) Control, different school	Public	H.S. Seniors	M + F	Ten verbal and ten math coaching sessions using ETS item materials	18.3	.01	161/158	6.2	.01	161/158
French (1955) Control, different school	Public	H.S. Seniors	M + F	Total verbal coaching 4½ hours, math coaching was ten sessions using ETS item materials	5.0	.05	110/158	18.0	.01	161/110
Dear (1958) Control, same and different schools	Public & Private	H.S. Seniors	M + F	Approximately 6 weekly 2-hour, 2-person coaching sessions, plus 1 hour of homework each week	-2.5	N.S.	60/526	21.5	.01	60/526
Dear (1958) Control, same and different schools	Public & Private	H.S. Seniors	M + F	Approximately 12 weekly 2-hour, 2-person coaching sessions plus 1 hour of homework each week	–[3]–	–[3]–	—	23.6	.01	71/116
Frankel (1960) Control, same school statistically matched	Public	H.S. Seniors	M + F	Ten 3-hour, 25-person sessions of coaching	8.4	N.S.	45/45	9.4	N.S.	45/45
Whitla (1962) Control, statistically matched	Public & Private	H.S. Seniors	M + F	Commercial proprietary school. Five 2-hour sessions plus intensive homework. Verbal and math	11.0	N.S.	52/52	-5.3	N.S.	50/50
Roberts & Oppenheim (1966) Randomized	Public	H.S. Juniors	M + F	7½ hours of programmed instruction in test-taking and in verbal and math content	14.4[4]	.05	154/111	8.1[4]	N.S.	188/122
Evans & Pike (1973) Randomized	Public	H.S. Juniors	M + F	Test-taking skills; math content. Seven 3-hour sessions, 21 additional hours of homework	No coaching for SAT-V		—	16.5	.05	288/129
Alderman & Powers (1979) Randomized	Public & Private	H.S. Juniors	M + F	Varied strategies, at eight schools, centering on reading and analogies; time range 5-45 hours	8.4	.05	239/320	No coaching for SAT-M		—
Average weighted by size of experimental sample					9.1			13.0		

[1] The coaching effects are intercept differences between regression lines for experimental and control groups or (for Frankel, Whitla, Roberts & Oppenheim, and Pike & Evans) average score increases of experimental over control groups, both weighted in the case of multiple experimental or control groups by their respective sample sizes.

[2] As shown for coaching effects reported in original text.

[3] Not calculated; variances and regression slopes differed significantly for experimental and control groups.

[4] This study employed the PSAT as both pre- and posttest; the averages shown have been converted to the SAT score-scale ranging from 200 to 800 points.

ies summarized in Table II-1 were 30 hours (plus homework) over 10 sessions (Frankel, 1960), 24 hours (plus 12 hours homework) over 12 weeks (Dear, 1958), and 21 hours (plus 21 hours homework) over 7 weeks (Evans & Pike, 1973).

In the absence of control groups, instructional or program effects were estimated in Table II-2 in the manner described earlier—that is, by adjusting the average score gains reported in each study by the average of four adjustments, those suggested by (1) the authors of the original articles, (2) Slack and Porter (1980), and (3) Pike (1978), as well as (4) the average score gains of control students in other coaching studies who had roughly comparable initial score levels. Averaging these estimates over all the studies in Table II-2, weighting in each case according to group size, yields 38 points for SAT-V and 54 points for SAT-M (the unweighted averages are 39 points for Verbal and 53 points for Math). Given the dubious and provisional nature of the adjustments and the highly self-selected character of the students in each program, these values are still probably overestimates of program effects But their general magnitude suggests that the verbal and mathematical reasoning skills measured by the SAT may be enhanced to a measurable degree by long-term and intensive instruction, at least for highly motivated students.

Granted that there is some overlap or blurring of the distinction between coaching studies in Table II-1 and instructional studies in Table II-2, the two types seemed sufficiently different to warrant separate treatment. Accordingly, overall averages were not computed for the total combined set of studies because possible differences in impact might thereby be obscured. In contrast, Slack and Porter (1980) have chosen to combine both types of studies in a single table and to report overall weighted average score increases of coached groups over uncoached control groups (or norm comparison groups); these weighted averages were 29 points for SAT-V and 33 points for SAT-M. However, their table included French's (1955) study of identical items, which hardly seems appropriate, and a study by Lass (1958) which we viewed as too informal to treat quantitatively, and it did not include the studies by Alderman and Powers (1979) or by Evans and Pike (1973). If the French identical-items study is deleted from their table and the two missing studies are added and some corrections are made in the effective sample sizes of the studies by Dyer (1955) and Pallone (1961), the revised weighted averages are 25 points for Verbal and 33 points for Math. Further, if the score increases taken from the studies by Pallone (1961) and Marron

Table II-2

Adjusted Average Score Gains in Studies of SAT Instructional Interventions Having No Control Groups

Study	*Sample Characteristics* *School*	*Level*	*Sex*	*Characteristics of the Special Preparation*	SAT-*Verbal Adjusted Average Score Inscrease*[1]	*N*	SAT-*Math Adjusted Average Score Increase*[1]	*N*
Pallone (1961)	Private	H.S. Seniors & Graduates	M	90-minute daily instruction and practice in developmental reading skills over 6 weeks	81	20 +	—	—
Pallone (1961)	Private	H.S. Seniors & Graduates	M	50-minute daily instruction and practice in developmental reading skills, with stress on logical inference and analogic analysis over 6 months	68	80 -	—	—
Marron (1965)	Private	H.S. Seniors & Graduates	M	Full-time daily sessions aimed at verbal and math content and test-taking skills over 6 months	Group 1 54 } 35	83	Group 1' 59	232
					Group 2 33	600	Group 2' 53	405
					Group 3 24	5	Group 3' 46	78
					Group 4 12	26		
Coffman & Parry (1967)[2]	Public	College Freshmen	M + F	6-hours weekly instruction in accelerated reading over 8 weeks	4[3]	19		
Weighted Average					38		54	

[1] To estimate instructional or program effects, average score increases in the Pallone (1961) and Marron (1965) studies were adjusted by the average of four adjustments: those suggested by (1) the authors of the original articles—Pallone suggested 35 points on SAT-V as normal expectation of gains during the final year of secondary school (15 points for the 5-month interval between tests in the short-term program), and Marron suggested 24 and 26 points, respectively, for SAT-V and -M as typical gains for high school seniors over 6 months; (2) Slack and Porter (1980)—average gains in national administrations of junior- to senior-year retesters having the same initial average score levels as Pallone's and Marron's groups; and (3) Pike (1978)—average gains of control students in superior schools from other studies of proprietary programs; as well as, (4) average gains of control students in other studies who have average initial score levels roughly comparable to Pallone's and Marron's groups.

[2] The 15-week program in Coffman and Parry (1967) was not included because the 29-point mean decrease in scores was considered atypical and possibly indicative of motivational and test administration problems.

[3] The two 8-week programs in Coffman and Parry (1967) were combined, but adjustments were made only by the Slack and Porter (1980) procedure, which attenuated by only a few points an already tenuous effect. None of the suggested comparison groups of SAT takers appeared to provide even remotely reasonable yardsticks for gauging score gains of students already enrolled in a college not requiring the SAT.

(1965) are adjusted by the average of four suggested control contrasts, as was done in this report's Table II-2, the adjusted weighted averages of Slack and Porter become 22 points for Verbal and 28 points for Math.

Thus, the overall effect of averaging together the two types of studies—those with control groups and those without—is to raise the estimates of average coaching effects based on control-group studies alone (Table II-1) by about 13 points on Verbal and 15 points on Math. This assumes that the adjustments applied to the score gains in studies lacking control groups were large enough to correct appropriately for the experiential growth of self-selected students that would have occurred regardless of the program. This combined averaging of Slack and Porter (1980) is misleading not only because the adjustments are questionable, but because the combined averages obscure important differences between the special preparation programs in the two types of studies. As indicated earlier, a comparison of the brief program descriptions in Tables II-1 and II-2 reveals that the programs in studies lacking control groups happened also to be quite long-term and intensive with respect to student contact time, while the programs in control-group studies were relatively short-term and nonintensive. The former programs also entailed organized curriculum content and skill development as well as test review, whereas the latter programs tended to emphasize test review and practice exercises.

Rather than averaging across these studies, which inevitably precipitates arguments about the appropriate size of the score effects to be included from uncontrolled studies, let us instead *rank* the studies in order of the reported treatment vs. control group contrasts and when control groups are not available in order of the reported score increases. That is, for SAT-V Pallone's (1961) long-term and summer programs would be ranked 1 and 2, respectively, followed by Marron's (1965) four groups, etc. This procedure grants that the score effects in the studies lacking control groups are larger in an ordinal sense than those in the control-group studies, but it takes no position with respect to how much larger. If the programs are then also ranked in terms of the number of student contact hours involved and a Spearman rank-order correlation coefficient computed (Table II-3), the rank correlation is found to be .60 across 22 studies for SAT-V and .80 across 11 studies for SAT-M. Both coefficients are significant at the .01 level (Dixon & Massey, 1951). If five particularly suspect studies are deleted from the calculations for SAT-V, the new correlation is .73 across 17 studies, which is also significant at the .01 level. In this

Table II-3

Correlations Between Rank Order of Score Effect and Rank Order of Student Contact Time for Studies of SAT Interventions With and Without Control Groups

Verbal[1]

Study	*Student Contact Time*	*Rank Time*	*Rank Score Effect*
Dyer (1953)	10 hours[2]	11	13
French (1955) V&M	8.3 hours[3]	12	7
French (1955) Vocab	4.5 hours	17	12
Dear (1958)	6 hours	13.5	16
Frankel (1960)	15 hours	8	11
Whitla (1962)	5 hours	15.5	10
Alderman & Powers (1979)[4]			
School C	10.5 hours	10	17
School E	6 hours	13.5	15
School F	5 hours	15.5	14
School G	11 hours	9	8
School H	45 hours	6.5	9
Pallone (1961)			
Short	45 hours	6.5	2
Long	100 hours	5	1
Marron (1965)			
Group 1	300 hours	2.5	3
Group 2	300 hours	2.5	4
Group 3	300 hours	2.5	5
Group 4	300 hours	2.5	6
Rank-Order Correlation		.73 (17 studies)	
Roberts & Oppenheim (1966)[5]	3.8 hours		
Alderman & Powers (1979)			
School A	7 hours		
School B	10 hours		
School D	10 hours		
Coffman & Parry (1967)[6]	48 hours		
Rank-Order Correlation		.60 (22 studies)	

Math[1]

Study	*Student Contact Time*	*Rank Time*	*Rank Score Effect*
Dyer (1953)	8.3 hours[7]	7.5	7
French (1955)[8]	8.3 hours	7.5	8
Dear (1958) Long	12 hours	6	4
Dear (1958) Short	6 hours	9	5
Frankel (1960)	15 hours	5	9
Whitla (1962)	5 hours	10	10
Evans & Pike (1973)	21 hours	4	6
Marron (1965)			
Group 1	300 hours	2	1
Group 2	300 hours	2	2
Group 3	300 hours	2	3
Rank-Order Correlation		.74 (10 studies)	
Roberts & Oppenheim (1966)[5]	3.8 hours		
Rank-Order Correlation		.80 (11 studies)	

[1] When only a total student contact time was available, it was assumed that half the time was devoted to Verbal and half to Math.

[2] Each of 12 exercises was estimated to require a 50-minute class period.

[3] Each of 10 exercises was estimated to require a 50-minute class period.

[4] Schools A, B, and D were suspect because of control group score decreases and hence were omitted in these calculations based on 17 studies; they are included below in calculations based on all 22 studies.

[5] Roberts & Oppenheim (1966) was suspect because of control-group decreases and hence was omitted from the calculations for Verbal based on 17 studies and for Math based on 10 studies.

[6] Coffman & Parry (1967) was suspect because of treatment-group score decreases and other indications of motivational problems and hence was omitted from the calculations based on 17 studies. The 25-week program was dropped altogether because of the treatment group score decreases, and the two 8-week programs were combined, yielding a weighted average score effect of 6.5.

[7] Each of 5 "double-length" exercises was estimated to require two 50-minute class periods.

[8] French's (1955) Math coaching group was contrasted with two control groups, one in a school having no coaching and one in a school having vocabulary coaching; these two comparisons were combined, yielding a weighted average score effect of 11.

latter calculation, Schools A, B, and D from Alderman and Powers (1979) were eliminated because of control-group score decreases, as was the Roberts and Oppenheim (1966) study for the same reason, and the Coffman and Parry (1967) study was dropped because of treatment-group score decreases and other indications of the low relevance to the SAT of both the accelerated reading program studied and the samples of enrolled college students employed. If one suspect study, namely Roberts and Oppenheim (1966), is deleted from the calculations for SAT-M, the new correlation is .74, which is also significant at the .01 level. These rank-order correlations are summarized in Table II-3. It should be noted that although the various coaching programs required different and unknown amounts of homework, this rank-correlation procedure tacitly assumes that the amount of homework in each case was roughly proportional to the number of student contact hours, so that the overall orderings would not be markedly changed if homework were taken into account.

In interpreting these sizable monotonic relationships between student contact time and score effects, it must be remembered that these are rank-order correlations between average values of different samples or groups, and correlations between averages are typically much higher than correlations between individual differences within groups. Furthermore, these rank correlations are dominated by the relative consistencies between the two types of studies with respect to the ranking variables—that is, the control-group studies are all relatively low in both student contact time and score effects while the uncontrolled studies are all relatively high in both student contact time and score effects. It must also be remembered that the relatively high-contact programs entailed structured curricula emphasizing content knowledge and skill development while the relatively low-contact programs emphasized test review and practice. With this confounding of program characteristics in mind, it appears that increases in student contact time (possibly serving as a proxy for increasing curriculum emphases on content knowledge and skill development) are systematically associated on the average with increases in SAT scores. However, even though the time dimension is covered in only a fragmentary fashion by the available studies, when the magnitude of (adjusted) score effect is plotted against student contact time, the relationship appears not to be linear but approximately logarithmic. There is somewhat more variance or noise around the line of best fit relating score effect to log contact time for SAT-V than SAT-M, but only the Pallone (1961) results appear to be particularly out of line in this logarithmic formulation. Considerably more research is needed, of course, with special preparation programs entailing more than 40 hours of student contact time to fill in the gaps on the time line. But, if this suggested logarithmic relationship has substance, then each additional increase in SAT scores may require geometrically increasing amounts of student contact time and of all the curricular effort that contact time may be proxy for.

In summary, the average coaching effect across studies having some type of control group was less than 10 points for SAT-Verbal and less than 15 points for SAT-Math. The special preparation programs in these control-group studies tended to be relatively short-term and relatively nonintensive in terms of student contact time, three of the longest and most intensive being 30 hours of student contact over ten sessions, 24 hours of student contact over 12 weeks, and 21 hours over seven weeks. For particular groups of students and particular coaching programs, the score increases ranged as high as about 18 points for Verbal and about 24 points for Math (or even somewhat higher if interactions between years of math taken and sex are considered). The larger score increases tended to be associated with the longer and more intensive of these relatively short-term programs, especially in Math. The average program effect across studies having no control groups could only be tentatively estimated, because there is no good way of taking into account the experiential growth of self-selected students in the absence of comparable control groups. The amount of their experiential growth and skill development could be considerable in this case, because all of the programs in the studies lacking control groups happened also to be relatively long-term. In addition, they were relatively intensive in terms of student contact time. The briefest of these programs was 45 hours of student contact over six weeks, and the longest was virtually full-time over six months. The provisional estimate of average program effects for these noncontrolled studies was 38 points for Verbal and 54 points for Math. Although the substantive content of the coaching programs was not systematically evaluated in any of these studies, overall the smaller coaching effects appear to be associated with short-term, relatively nonintensive practice and review, and the larger effects (which occur more for Math than for Verbal) appear to be associated with longer-term, high student-contact programs focussing on skill development.

III

Beyond the FTC Study: Critique of Assumptions and Inferences

This section draws together the major points made in several reviews undertaken at Educational Testing Service (ETS) of the Federal Trade Commission's (FTC) study of the effectiveness of commercial coaching schools in raising scores on the College Board Scholastic Aptitude Test (SAT). The separate reviews overlap somewhat in their criticisms, but they are reproduced in full in the Appendix since some additional points not covered in this summary are raised. The comments that follow are limited to the revised statistical analyses issued by the FTC Bureau of Consumer Protection (BCP) in March of 1979, undertaken in response to the several major data-analysis flaws observed in the memorandum issued previously by the FTC Boston Regional Office (1978).

Before proceeding to develop the critical arguments, we will first briefly describe the object of this critique, namely the FTC study of the effects of commercial coaching on the SAT. Students enrolled in two New York City area commercial coaching schools during the testing years 1974-75, 1975-76, and 1976-77 served as the experimental or treatment group, and a random sample of uncoached persons who took the SAT during the same three-year period in the same greater New York metropolitan area served as a control group. Data from a third coaching school were not analyzed because of its small sample size. Six subgroups were analyzed: (1) high school juniors taking the SAT for the first time in April 1975 (76 coached and 607 uncoached students); (2) juniors taking the SAT for the first time in April 1976 (247 coached, 617 uncoached); (3) seniors taking the SAT for the second time in November 1975 (98 coached, 396 uncoached); (4) seniors taking the SAT for the second time in November 1976 (177 coached, 387 uncoached); (5) all high school students taking the SAT for the first time on any test date during the three-year period (417 coached, 1763 uncoached); and, (6) all high school students taking the SAT for the second time during this period (316 coached, 1267 uncoached). Statistical analyses were actually based on smaller samples than these largely because of missing student descriptive data.

Since the coached and uncoached groups might differ from each other in a number of ways possibly relevant to the treatment, the demographic and personal characteristics of the two groups were contrasted. It was found that the coached group was significantly higher than the uncoached group in high school class rank, parental income, most recent English grades, most recent math grades, and number of years of math taken; in addition, the coached group included significantly more nonpublic school students and fewer public school students than the uncoached group. Before multiple regression analyses controlling for these and other background variables were conducted, the possibility of differential coaching impact on good and poor students was first discounted by noting a lack of interaction between PSAT scores and coaching treatment.

The multiple regression analyses, which controlled for PSAT (or first SAT) as well as for the several relevant background variables, yielded neglible effects for students at one school and statistically significant effects for students at the other, where the impact for SAT-Verbal was found to be 30 and 27 points, respectively, for first- and second-time SAT takers over the pooled time periods and 19 and 28 points, respectively, for SAT-Math over the same periods. Since these values represent combined coaching and self-selection effects by virtue of the confounding between pre-existing group differences and the coaching treatment, the FTC report then presented an analysis of potential self-selection bias. In a regression analysis of the pooled treatment and control groups, coached students were found to achieve lower PSAT (or first SAT) scores than were predicted from their background characteristics, whereas uncoached students scored slightly higher than expected given their personal and demographic background. This type of self-selection was characteristic of students attending the apparently effective coaching school but was not generally found at the ineffective school. In an effort to eliminate the self-selection effect of this underachievement on the PSAT (or first SAT), the regression analyses were repeated dropping the PSAT (or first SAT) from the set of covariates. As a consequence, the estimated effects were greatly reduced. For example, the effects for SAT-V and -M for first-time SAT takers over the pooled time periods in the previously effective school were 11.5 and .55, respectively, which are no longer statistically significant; the effects for second-time SAT takers over the same period were 16.2 and 16.6, respectively, for SAT-V and -M, which remain statistically significant. It was then argued that this reanalysis would be appropriate only if the underachievement on the PSAT for coached students were due to chance. Since an analysis of PSAT scores for students coached between their first and second SAT exams revealed a similar pattern of underachievement on both the PSAT and the first SAT, it was concluded that the phenomenon was not random but was characteristic of students self-selecting coaching, that is, they were underachievers on standardized tests and would likely continue to be test underachievers in the absence of coaching. Therefore, it was argued, the prior results showing coaching effects in the 20- to 30-point range for both SAT-V and -M at one coaching school were the most defensible findings. Those findings still represent estimates of combined coaching and self-selection effects, however, since this attempt to analyze self-selection as underachievement on standardized tests did not alter the confounding of coaching with pre-existing group differences or otherwise eliminate the effects of unmeasured self-selection factors.

The most fundamental issue concerning the FTC coaching study is that it undertakes an interpretation of *available data* as a substitute for collecting *experimental data*. The fact that the data are taken from records and files necessarily puts the study out of reach of the kinds of experimental controls that would permit clear, unambiguous interpretation of findings—i.e., it is a quasi-experiment rather than a true randomized experiment. Such a study does not involve random assignment of students to coaching and noncoaching conditions, and in the absence of randomization some interpretive equivocality is inevitable. We hope to reduce—but cannot eliminate—this equivocality by conducting multiple alternative statistical analyses. Summaries of two such reanalyses are included in Sections IV and V of this report and the full texts are reproduced in the Appendix.

The Power of Randomized Experiments

The value of a randomized experiment warrants discussion here. To conduct a true experimental study of coaching, one assembles a large group of students representative of the kinds of individuals about whom inferences and generalizations are to be drawn. These students are assigned at random to either a treatment (or experimental) subgroup to receive coaching or to a control subgroup for whom the coaching experience is to be delayed. To increase precision, before the treatment subgroup takes the coaching course, a form of the SAT could be administered to both subgroups as a pretest. Or, to avoid possible pretest-treatment interactions, a different instrument might be used as a pretest or some other proxy measure of ability or achievement used as a covariate. Effective control conditions should be established to maintain motivation, avoid attrition, and otherwise assure that the two groups remain comparable except that one receives coaching and the other does not. At the end of the coaching period, the SAT is administered as a posttest to both treatment and control subgroups. The experimental data may be analyzed in any one or a combination of several ways. For example, the analysis-of-covariance model uses pretest and other pretreatment variables to adjust for any differences that might exist between the two randomly assigned subgroups on those variables. Since by randomized design and the maintenance of effective controls the only systematic difference between the two subgroups is that one received coaching and the other did not, differences observed on the outcome or posttest measure can be confidently attributed to the coaching experience within some range of standard error.

In a nonrandomized design or quasi-experiment such as the FTC study, in which the coached students were those who had enrolled in coaching schools and the "control" students were drawn from another source, there is no way of discounting alternative reasons for the difference observed on the outcome or posttest measure. The difference might result from the coaching experience or, on the other hand, might simply reflect differences in the characteristics of the two groups existing prior to the coaching experience. One powerful feature of the randomized experiment in this regard is that we can attach probabilities to the likelihood of these alternative events. Thus, although we have other reservations regarding the analysis and interpretation of the FTC study, our chief reservations are those that relate in one way or another to the nonrandomized nature of the study.

Confronting Treatment and Control Group Differences

As a consequence of the nonrandom assignment of subjects in the FTC study—indeed, the highly self-selected nature of the treatment group—the data offer numerous opportunities for alternative interpretations. The FTC-BCP (1979) report recognizes the limitations of quasi-experimental data and takes care to avoid the type of simplistic analyses undertaken by the Boston Regional

Office (1978). The BCP analysis takes proper note of the fact that the coached and uncoached groups differ at the one percent level of significance on several key demographic variables relevant to SAT performance, e.g., class rank, parental income, ethnic background, high school type (public or nonpublic), most recent English grade, most recent math grade, and expected years of math.

It is noted in this regard that the BCP analysts took account of certain demographic items available to them from the Student Descriptive Questionnaire (SDQ), but excluded others that seem at least as relevant as those that were included. For example, the SDQ data provided by the College Board and ETS for the study included such variables as students' level of educational aspirations and their parents' level of education, but these variables were omitted on the master tape released by BCP for reanalysis.

There are, in addition, some refinements that might have been introduced in the BCP analysis, but were not:

- It might have proven useful to conduct additional analyses of interactive effects—e.g., the coaching effect with sex of student, with minority status, with the number of previous PSATS taken, the number of years of math study, and high school grades. Analyses to assess possible interactions with treatment have been undertaken by ETS.

- In comparing coached and uncoached students, it would have been informative to report the demographic profiles for each coaching school separately rather than combining them across schools and risking the chance that important school effects might be obscured.

- Although the primary objective of adjusting for the independent variables was fulfilled, it might have been more informative if these variables had been entered into the analysis seriatim rather than simultaneously. This would have made it possible to examine the incremental effects of each variable as it entered into the regression system.

The Bane of Self-selection

We have some major questions regarding the formulation of some of the hypotheses and inferences drawn in the FTC report about the nature and impact of self-selection:

- The FTC report suggests that the control group "may in fact have received some form of coaching other than formal enrollment in a commercial coaching course" (p. 2), in which case the apparent benefit from coaching, as observed in this study, would have been an underestimate of the real coaching effect. However, data from a study by Powers and Alderman (1979) on the effects of using the College Board booklet *Taking the SAT* suggest that coached students may be more likely than uncoached students to engage in additional methods of preparation. Thus, under these circumstances the FTC analyses would be more likely to overestimate the effects of commercial coaching schools by failing to account for other ways in which coached students may have prepared for the test.

- The hypothesis is offered in the FTC report that students who are coached before taking the SAT for the first time are serious about their test experience and *do not* plan to take it a second time, while those who are uncoached are not as serious since they *do* plan to take the test a second time. If this is so, then higher scores obtained by coached students may not be due to coaching but to the difference in how serious the students are. The results showed that coached students were, in fact, not less likely but more likely than uncoached students to take the SAT a second time and the conclusion is drawn that the issue of seriousness is, therefore, not a cause for concern. However, consideration should be given to the counter-hypothesis that because coached students were more likely than uncoached students to take the SAT a second time, coached students are indeed *more* serious and *more* determined to persist until they obtain satisfactory SAT scores.

- In considering the possibility of self-selection, PSAT (or SAT) scores were predicted from a composite of demographic characteristics including family income, most recent English and mathematics grades, and class rank, among others, for the pooled sample of coached and uncoached students. These predicted scores were then compared with actual initial PSAT (or SAT) scores. The coached students were found to earn lower initial PSAT (or SAT) scores than expected from their background characteristics, whereas uncoached students, on balance, scored slightly higher than expected given their personal and demographic characteristics. Thus, the coached students on the average achieved lower initial PSAT (or SAT) scores than uncoached students having the same values of demographic variables. It is concluded that coached students "tend to underachieve on standardized tests" and, further, that coaching is "effective for underachievers." The terminology is rather unfortunate in that it is easy to slip, as happens occasionally in the FTC report, and refer to the coached students as "underachievers"—a term that is generally applied to students whose grades in school are lower than one would expect on the basis of test scores or prior achievement. In fact, the comparison of demographic profiles presented for coached and uncoached students indicates that the coached group earned higher grades in English and mathematics and had higher class ranks than did the uncoached students—but they did not score as well on the PSAT as those accomplishments would have led one to expect.

In contrast to the hypothesis offered in the FTC report that "the students who choose to go to a coaching school represent those individuals who scored lower than they expected on the SAT" (p. 23), a number of other hypotheses about self-selection are at least as plausible and consistent with the data. These include self-selection on the basis of family income, attending private schools, etc. The point is that we simply cannot tell with any certainty on what basis self-selection occurred. For example, students at Coaching School A, which charged $225 for the course, were found to be "underachievers on standardized tests," whereas students at coaching School B, which charged only $75, were not.

- In the FTC regression analyses, the estimated effects of taking the PSAT twice prior to taking the SAT were found to be larger, although less precise, than those attributed to coaching. These effects on SAT performance of repeating the PSAT are unusually high relative to what we know about the impact of prior test practice or test repetition, suggesting the operation of some form of self-selection bias in these data.

- A major conclusion of the study is that "coaching can be effective for those who do not score well on standardized tests" (p.35). On closer examination, we note that this conclusion is not substantiated in data analysis because the FTC regression analyses, by virtue of not investigating interactions with treatment, estimate the same effect of coaching regardless of demographic characteristics of the subjects. Consequently, there is no evidence in the FTC study on this point of differential effectiveness of coaching. Rather, this conclusion is based on conjecture, as follows: One school produced significant coaching effects and the other did not; students at the effective School A tended to be underachievers on standardized tests while those at the ineffective School B were not. The FTC report then notes that coaching can be effective for underachievers on standardized tests because it was at School A, rather than arguing that underachievement on standardized tests and attendance at School A are confounded and that self-selection affords a plausible rival interpretation for the obtained effects or for part of those effects. It is then speculated that "if only underachievers can be helped, it is possible that coaching at School B would be effective for such students" (p. 36). Furthermore, in School A self-selection in the sense of "underachievement on standardized tests" was not apparent on one of the test administrations (the second SAT-Verbal taken in 1975), and the conclusion is drawn that "coaching at School A can be effective for all students, not just for underachievers" (p. 35). Although qualified in the FTC report, for conclusions of this importance to be drawn with such generality one would expect both the data and the analyses to be more directly supportive.

Additional Puzzlements

The FTC report describes some findings that warrant clarification and further research:

- The effects of coaching on the Verbal section of the SAT were found to be as large or larger than the effects on the Math section. This result seems contrary both to intuition—the Math test is much more curriculum-related than is the Verbal test—and to the general trend of results from earlier studies that the Verbal test is less susceptible than is the Math test to special preparation effects.
- The FTC report gives data showing that the effect for students attending Coaching School A was significantly greater on SAT-Verbal in 1976 than in 1975 and greater on SAT-Math for second-time takers than for first-time takers. These variations are potentially important, and an examination of their sources might illuminate the relative effectiveness of coaching practices.
- The analysis revealed that attendance at School A was associated with larger effects on SAT performance than attendance at

School B. Considering the fact that the tuition at School A is three times the tuition at School B, it is not surprising that the self-selection effects discussed in the report are stronger for the School A students than for the School B students, who might thereby be expected to be more similar to the uncoached students. Such a view would put more stress on the School B findings than is given in the FTC report.

At a more specific level, there is some concern that the procedures used for constructing the samples in the FTC analyses may have introduced biases related to the effects under investigation. For example, the FTC noted that testing histories were identified for 1568 of the coaching school students but not for 600 others. This represents an attrition of about 28 percent, a highly significant loss. In addition, all students who did not respond to the Student Descriptive Questionnaire were dropped from the analysis. The study showed a significantly lower rate of response to the items on the SDQ—approximately 64 and 67 percent for the coached and uncoached students, respectively—than is typical of the College Board candidate group. The usual response rate for the candidate group is about 86 percent.

New Perspectives on the FTC Data

In an attempt to clarify some of the points raised in this critique, ETS undertook reanalyses of the data included on the master tape released to the public by the FTC. As part of our effort to understand the conduct and implications of the FTC study, ETS invited Dr. T.W.F. Stroud, Queens University, Kingston, Ontario, to design and carry out additional analyses of the FTC master tape data. As the first step in his independent reanalysis, Dr. Stroud successfully replicated the results as reported by the FTC. Having assured the accuracy of the FTC computations using their statistical design, Dr. Stroud undertook a reanalysis of the data for all three coaching schools using a similar analysis-of-covariance model, but differing from the FTC analysis in two important respects: The FTC analysis employed a pooled regression equation across both coached and uncoached students using, along with other variables, PSAT-V in predicting coaching effects on the SAT-Verbal score and PSAT-M in predicting SAT-Math effects. In contrast, Dr. Stroud employed a regression equation defined by the performance of uncoached students only and, in addition, included both PSAT-V and -M in predicting coaching effects for each area on the SAT. This analysis is preferable not only because it is more precisely controlled through the inclusion of additional covariates, but because it more appropriately contrasts the performance of coached students with predicted levels based on uncoached students rather than a mixture of the two. This approach results in valid estimates with fewer assumptions and enables examination of interactions in a straightforward manner (Cochran, 1968).

This reanalysis indicates that, given their background characteristics and pretest levels on the PSAT, students enrolled in one of the three coaching schools studied obtained significantly higher SAT scores than did uncoached students by about 20 to 35 points in both Verbal and Math—the same neighborhood as the FTC estimates. The estimated effects of coaching for the other two schools were not statistically significant. These are estimated combined effects due to coaching and self-selection (self-selection in terms of students who seek and complete commercial coaching programs as opposed to those students who do not), since it is not possible to estimate coaching and self-selection effects separately with these data. No interactions were uncovered at Coaching School A for either SAT-V or SAT-M. At Coaching School B, however, statistically significant and independent interactive effects were obtained on SAT-V for race and self-reported parental income. On the average, even though their number was quite small (N = 13), black students at School B exhibited significantly larger coaching/self-selection effects on SAT-V than nonblacks, and students reporting low family income exhibited significantly larger coaching/self-selection Verbal effects than those reporting high family income. A detailed summary of the Stroud reanalysis follows in Section IV, and the complete report appears in Appendix 2.

A second study of the data was undertaken by Donald A. Rock, Senior Research Psychologist at ETS. Since we know that gains in SAT scores can be expected from the junior year to the senior year of high school, Dr. Rock applied a statistical model incorporating growth effects. For the treatment or experimental group, this study included only those students at the largest coaching school (School A) for whom three sets of test scores were available, a PSAT and two administrations of the SAT. The control group included only those uncoached students for whom these same three sets of test scores were also available. The coached students, who were labeled "underachievers" in the FTC report, performed better than the uncoached students on the PSAT and were also higher in high school rank-in-class and family income than were the uncoached students. As would be expected, the coached students also scored higher than their uncoached cohorts on the first SAT administration, but in the case of the Verbal area scored differentially higher. That is, during the period from taking the PSAT to taking the initial SAT, prior to attendance at the coaching school, the verbal skills of the coached students appeared to grow more rapidly than those of the uncoached students. In Math, however, both the coached and uncoached students showed similar group growth rates during this preintervention period. When the confounding effects of differential group growth rates are controlled for, the estimated coaching effect on the Verbal score (about 17 points) is substantially smaller than that for the Math score (about 30 points). The fact that these two effects are different from each other, and not the same, is consistent with the results of earlier studies and with expectations that Math, being generally more curriculum-related than Verbal, might be more responsive to special preparation. A detailed summary of the Rock study follows in Section V, and the complete report appears in Appendix 3.

Most of the criticism in this section points up the limitations of the FTC data and the limits on inferences that can be drawn from such data. Within these limits, further analysis and interpretation may still prove to be illuminating. In this spirit, we next attempt to estimate the combined coaching/self-selection effects using more refined methods and to investigate the possibility of interactions between size of effects and student background characteristics. We also attempt to adjust statistically for that portion of self-selection effects that is embodied in differential group growth rates.

IV

Estimation of Combined Coaching / Self-Selection Effects in the FTC Study: Detailed Summary and Elaboration of the Stroud Reanalyses

In an effort to obtain more precise estimates of the combined coaching/self-selection effects in the FTC data, Dr. T.W.F. Stroud of Queens University, Kingston, Ontario, at the request of ETS, designed and conducted additional analyses introducing several refinements over the procedures used by the FTC. It will be recalled that the FTC study was based on six subsamples: (1) high school juniors taking the SAT for the first time in April 1975 or (2) in April 1976; (3) high school seniors taking the SAT for the second time in November 1975 or (4) in November 1976; and, (5) all high school students taking the SAT for the first time on any test date over the three-year period 1974-1977 or (6) taking the SAT for the second time on any test date during that period. Data for students at two coaching schools in the metropolitan New York area were included, while data from a third coaching school were left unanalyzed because the number of students was considered too small. A control sample of uncoached students consisted of every 150th individual in the ETS files who took the SAT during the given three-year period in the same greater New York area.

The following background variables, roughly in decreasing order of their importance, were controlled for in the FTC regression analyses: pretest score [PSAT-V when the first SAT-V (SAT1-V) was being predicted, SAT1-V when the second SAT-V (SAT2-V) was being predicted, and similarly PSAT-M when predicting SAT1-M and SAT1-M when predicting SAT2-M], self-reported grade in English or Math, self-reported rank-in-class, self-reported years of English or math taken, self-reported parental income, sex, race, high school type (public or nonpublic), and number of PSATS taken. Time between pretest and posttest was also included as a covariate, but it did not improve prediction significantly. The FTC regression analyses were based only on students with complete data on all of these variables. For each of the six subsamples, separate regression equations predicting SAT-V and SAT-M were computed for pooled coached and noncoached students, using dummy variables to represent attendance at Coaching School A and Coaching School B.

Overall, the FTC analyses showed that students at Coaching School A, on the average, scored significantly higher on the SAT than noncoached students. The amount of advantage on a 200- to 800-point score scale falls somewhere between 14 and 38 points, which are 95% confidence limits for the differences in adjusted means (the median lower confidence limit and the median upper confidence limit over the 12 analyses, for SAT-V and -M in the 6 subsamples). Students at Coaching School B did not do significantly better than noncoached students; their score effect falls somewhere between median confidence limits of -12 and +19 points. In any event, since the coached students in this nonrandomized study differed significantly from the noncoached students in a number of ways that could have influenced both their decision to attend coaching school and their performance on the SAT (such as having higher rank-in-class and higher parental income), these obtained score differences between coached and uncoached groups represent a confounding of coaching effects and personal factors that cannot be disentangled with the available data. Among these personal factors influencing attendance at commercial coaching schools, for example, are motivation to earn a higher test score and financial means. The FTC data set includes a rough proxy for financial means in the form of self-reported parental income, but it includes no proxy for motivation or for a host of other important ways in which coached and noncoached students are likely, on the average, to differ, such as in career aspirations or in level of parental education. Attempts were made in the regression analyses to control for student differences in reported income, but there is no way that statistical adjustments can take unmeasured influences into account. As a consequence, the score effects reported in the FTC study must be interpreted as combined coaching/self-selection effects, as must the findings of the reanalyses that follow.

The reanalyses undertaken here differ from the FTC analyses in a number of respects:

In the FTC analyses, verbal pretests (PSAT-V and SAT1-V) and verbal background variables (e.g., grades in English, years of English) were used to predict SAT-Verbal scores, and quantitative pretests and background variables were used to predict SAT-Math scores. However, since the inclusion of both verbal and quantitative variables improved prediction of each score, both verbal and quantitative pretests and background variables were used in the current reanalyses to predict both SAT1-V and -M for juniors and both SAT2-V and -M for seniors. The only exception was "number of years of English," which was dropped from the analyses because it did not add to prediction in any regression equation when the other variables were already entered.

The current reanalyses dealt only with the peak test dates in 1975 and 1976 (subsamples 1, 2, 3, and 4 in the FTC study). Subsamples 5 and 6 for the pooled time periods were omitted because of their heterogeneity.

Only students with complete data were included in the FTC analyses. In contrast, Stroud used missing-value techniques so that students not reporting parental income, race, or rank-in-class could nevertheless be meaningfully included in the analyses.

In the FTC study, students enrolled in coaching school who did not receive coaching prior to the SAT administration in question were added to the uncoached group. Since the representativeness of the control sample is thereby eroded, these students were excluded from the present analyses.

All three coaching schools in the FTC data set were included in the present analyses, and effects and standard errors are estimated for each school separately. In addition, smoothed estimates are provided for the three schools which utilize the empirical Bayes concept of "borrowing strength" across samples and which allow, under certain assumptions, for the possibility of predicting coaching/self-selection effects in the same schools in future years.

In a key departure from the FTC methodology, a multiple regression equation predicting each dependent variable (SAT1-V and -M for juniors and SAT2-V and -M for seniors) is constructed in the present approach on the respective junior and senior samples of uncoached students rather than on a pooled sample of coached and uncoached students as in the FTC study (Belson, 1956). This procedure yields an unbiased estimate of the treatment effect in the presence of interactions between the size of effects and values of a covariate; it is also the recommended procedure when the control sample is much larger than the treatment sample (provided that the relative contribution of sampling errors in the control-group regression line to the variance of the effect estimate is negligible) (Cochran, 1968). These regression equations are then applied to the *coached* students to predict the SAT scores they would have expected had they been *uncoached* students with their same values on predictor variables. Since we wish to assess the average score increase of those *coached*, we adjust the *control* group from one mean to the other. Residuals from these regressions are then calculated for students at each coaching school, and the average value of these residuals is taken to represent the combined coaching/self-selection effect for each school. Combined coaching/self-selection effects, together with standard errors (including the sampling errors of the regression line), are reported in Table IV-1 separately for juniors and seniors in each of the two administration years for each of the three coaching schools. The overall results are similar to the FTC findings, although more precisely estimated with smaller standard errors. Combined coaching/self-selection effects were statistically significant for one coaching school and negligible and inconsistent for the other two coaching schools. The significant effects ranged from about 27 to 34 points for SAT-V and from about 20 to 33 points for SAT-M. Smoothed estimates based upon empirical Bayes procedures present a similar pattern in Table IV-2, with the size of the effects somewhat moderated and the standard errors somewhat increased.

Table IV-1

Coaching/Self-Selection Effects and Their Standard Errors

	SAT-VERBAL					SAT-MATH				
	1975 SAT1-V			1976 SAT1-V		1975 SAT1-M			1976 SAT1-M	
	School	*Effect*	*S.E.*	*Effect*	*S.E.*	*School*	*Effect*	*S.E.*	*Effect*	*S.E.*
Juniors	A	27.44	6.66	34.13	5.23	A	22.36	7.45	20.37	5.85
	B	—	—	4.16	6.38	B	—	—	3.71	7.14
	C	17.88	11.19	25.64	14.10	C	27.68	12.52	-23.32	15.79
	1975 SAT2-V			1976 SAT2-V		1975 SAT2-M			1976 SAT2-M	
	School	*Effect*	*S.E.*	*Effect*	*S.E.*	*School*	*Effect*	*S.E.*	*Effect*	*S.E.*
Seniors	A	32.07	5.86	31.34	4.51	A	22.54	6.41	32.62	4.93
	B	11.12	10.16	4.76	8.05	B	26.40	11.10	8.07	8.80
	C	-10.47	14.02	20.30	18.70	C	24.34	15.33	-21.57	20.44

Another advantage of the present methodology, in addition to yielding valid estimates with fewer assumptions and usually greater precision than the FTC approach, is the straightforward manner in which potential interactions may be explored. This is accomplished in general by correlating the coached students' residual scores, which represent individual coaching/self-selection effects, with whatever student characteristic or background variable one suspects might be associated with differential effects. Variables yielding significant correlations then serve to identify

Table IV-2

Smoothed Estimates of the Mean Coaching/Self-Selection Effects at the Three Coaching Schools

	SAT-VERBAL			SAT-MATH		
	School	*Effect*	*S.E.*	*School*	*Effect*	*S.E.*
Juniors (SAT1)	A	28.5	9.2	A	16.9	18.6
	B	9.7	12.8	B	7.8	23.0
	C	20.9	10.7	C	8.4	19.5
	School	*Effect*	*S.E.*	*School*	*Effect*	*S.E.*
Seniors (SAT2)	A	28.4	12.1	A	24.0	15.1
	B	9.4	12.3	B	17.2	15.3
	C	6.1	15.0	C	12.0	17.6

subsets of the coached student population that systematically exhibit greater coaching/self-selection effects than others. As previously suggested, it would not be surprising to find a number of interactions in the FTC data for such variables as "years of math taken," which revealed differential score effects for Dyer (1953)—possibly moderated by sex as it was for French (1955)—or educational or cultural disadvantagement as anticipated but not confirmed by Roberts and Oppenheim (1966).

In the present exploratory analysis of interactions, as an overall test, two multiple correlations were computed for the coached students, one predicitng the individual Verbal coaching/self-selection effects and the other the individual Math effects, using as predictors the same set of variables that had served as control measures in the initial regression equations computed on uncoached students. In the uncoached group, of course, these variables would be uncorrelated with residual scores. These predictors included PSAT-V and -M (or SAT1-V and -M), high school type, rank-in-class, years of mathematics, self-reported parental income, grades in English and math, sex, and race or minority group membership. (Two minority grouping variables were included: Group 1 refers to Asian Americans, Hispanics, and other nonblack minorities; Group 2 refers to blacks). Only persons with complete data from the three largest cells of coached students were included in this analysis, namely the 1976 juniors and seniors of School A and the 1976 juniors of School B; the sample sizes were 103, 102, and 85, respectively. For the two School A samples, neither the multiple correlation for the Verbal effects nor the multiple correlation for the Math effects was statistically significant. For the School B juniors, the multiple correlation for the Math effects was also nonsignificant but that for the Verbal

977

effects was significant at the .05 level. An examination of the individual correlation coefficients between each predictor and the Verbal residual score revealed that Group 2 (black vs. nonblack) was the most significant predictor variable. The most important predictor after Group 2 was entered into the equation was self-reported parental income. Group 2 correlated .288 with the Verbal residual score and Income correlated -.250, both of which are statistically significant at the .05 level. The partial correlation coefficient between Income and Verbal residual score after partialling out Group 2 was -.258, indicating that the effect of income and the effect of black vs. nonblack were essentially separate from each other in this sample of School B juniors.

The interaction with Income indicates that among coached juniors at School B those students reporting lower parental incomes exhibited larger coaching/self-selection effects on SAT-Verbal than did students reporting higher parental incomes. Since students with low family income may have had less access to learning resources at either home or school, this interaction might have arisen because some of them chose to attend coaching school as a limited-cost effort to gain short-term intensive access to such resources, which they then used to good compensatory effect. Or, as we have previously suggested, since both motivation and financial means are probably instrumental in deciding whether or not to attend a commercial coaching school, it is not unreasonable that coaching school enrollees having lower than average financial means might have especially strong motivation to perform well. It should be noted, however, that no interaction with parental income was uncovered for either juniors or seniors at School A, where the required financial investment is three times that of School B. The reason for such apparent differences in the effects of parental income in School B as opposed to School A is not obvious, unless the nonreplications signify that this interaction is one of those rare random events that achieves statistical significance by chance. On the other hand, the range of income was so restricted at School A—50 percent of the students reported parental incomes in the top category of more than $30,000 a year—that a significant correlation with residual scores would be difficult to obtain.

Plausible interpretations for the interaction with black vs. nonblack are even less obvious. For example, at first glance one might expect this interaction to be associated with disadvantagement. But if so, it is unrelated to economic disadvantagement because the interactive effects of black vs. nonblack in School B are separate from the effects of self-reported parental income; in this sample, unlike the general population, these two variables are uncorrelated. Nor is educational disadvantagement or developed ability level implicated in any simple way, because there were no differential effects associated with pretest score level or with high school grades. Neither was this Group 2 interaction replicated in School A, but in this case it had no opportunity to occur there because only three black students with the requisite background scores were identified in the School A data set.

In an effort to clarify the nature of this Group 2 interaction, test score and background data were examined for all of the coached students who identified themselves as black and who took the SAT either as juniors or seniors during the peak month of either administration year and for whom the information necessary for the missing-value regression equations was available. Fifteen such students were identified, 13 of whom attended Coaching School B. The average Verbal coaching/self-selection effect for these 13 students was 46.7 points above that for white students at School B, which is statistically significant at the .001 level. The average Math effect for these 13 black students also exceeded that for whites, but the difference was not statistically significant. The Verbal effects for the other two black students exceeded the cell mean for their respective coaching school by 24.6 points on the average. Six additional black students were identified who had taken the SAT at nonpeak administrations; their Verbal effects exceeded their corresponding cell means by 52.6 points on the average.

In terms of background characteristics, the total group of black students identified reported parental incomes ranging from 5 thousand to more than 30 thousand dollars per year, with a mean of $18,500. These students, in the main, were in the top third of their high school class and earned grades of B or better in both math and English, but their scores on the PSAT or SAT1 were relatively low. This pattern of lower test scores than might have been expected from background characteristics is reminiscent of the type of self-selection attributed to School A students in the FTC report, but such "underachievement on standardized tests," it will be recalled, was not found to be typical of School B students on the average. Since the Group 2 interaction might thus be another example of self-selection effects, data were examined for coached and uncoached black students in the Alderman and Powers (1979) randomized study of high school special preparation programs. Only 8 coached black students were identified but their individual SAT-Verbal coaching effects were estimated anyway, especially since the interaction in question was based on only 13 black students. Using the same within-school regression equations employed by Alderman and Powers, the black students' coaching effects were found to exceed their respective school mean coaching effect by only 8.4 points on the average, indicating little differential effectiveness for black students. Using within-school regression equations based only on uncoached white students, as a parallel to the regressions Stroud used for uncoached students in general, the coached black students scored an average of 3.8 points below the level predicted had they been uncoached whites. In contrast, the coached white students scored an average of 9.2 points above what would be expected had they been uncoached. Eight uncoached black students scored an average of 18.2 points below their expectation had they been uncoached whites with their same values on the predictor variables (in this case PSAT-V, PSAT-M, and Test of Standard Written English). The average residual score for the coached black students was thus 14 points higher (-3.8 vs. -18.2) than that of the uncoached blacks, while the average residual for the coached white students was 9 points higher than the expected value for uncoached whites. Hence there is little indication that coaching, at least in the form of high school special preparation programs as opposed to commercial coaching, works differentially better for black students in the absence of self-selection effects.

Whatever the basis for this black vs. nonblack interaction, it does not appear to be minority status in general, because Group 1 (Asian Americans, Hispanics, and other non-black minorities) did not yield a significant interaction. Indeed, when data for some 40 nonblack minority students were examined, their individual coaching/self-selection effects bore more resemblance to those of white students than black students, with the Verbal effect being only about 8 points higher than the corresponding cell means on the average. Tending to elevate that average, however, were 8 Asian American students, 4 at School A and 4 at School B, who exhibited Verbal effects about 46 points greater than their cell means on the average.

One possibility is that the Verbal SAT entails additional dimensions of difficulty for blacks, and possibly for Asian Americans as well, that might be overcome by special preparation—similar perhaps to the test sophistication factor found by Vernon (1962) in the multiple-choice responses of British students relatively unfamiliar with such tests, but not in the responses of American students. (This would be particularly plausible if some of the black students in the FTC sample were nonindigenous to the American educational system—from the West Indies, for example.) Countering this hypothesis, however, are the findings of Rock and Werts (1979), who compared factor structures for SAT-V and -M across Native American, Mexican American, Asian American, Puerto Rican, black, and white samples and evaluated the degree to which the same factors were being measured in each population. They concluded that both SAT-V and -M were measuring the same things across these populations, in the same units of measurement, with equal accuracy as indicated by equivalent standard errors of measurement. Moreover, on the average, SAT-V correlates equally well with college performance for whites and blacks, as does SAT-M (Breland, 1979).

Nonetheless, differences in test-taking strategies and in test sophistication more generally might still be at the root of the Group 2 interaction at Coaching School B, but it is difficult to limit that interpretation to black vs. nonblack differences per se. For example, in low-income, inner-city samples that were 75 percent black, Flaugher and Pike (1970) found more random responding on the PSAT than appeared in national administration samples as well as a tendency for the more able students to omit more items, which is counter to the usual trend. When patterns of guessing and omissions on the Graduate Record Examination were contrasted for black and white samples matched on total score level, however, no differences in the response patterns of black and white students were obtained (Pike, 1980), suggesting that the previously observed patterns might possibly be characteristic of low-scoring students in general rather than being specific to inner-city blacks.

In sum, the meaning of the Group 2 interaction for SAT-V at Coaching School B is both intriguing and perplexing. Even though it is based on only 13 students, the size of the effect was quite large and highly significant statistically. It is clear that well-controlled *randomized* studies of the possible differential effectiveness of coaching or special preparation for the SAT are needed to clarify this issue and to draw implications for practice.

V

Disentangling Coaching Effects and Differential Growth: Detailed Summary of the Rock Analyses

To recapitulate briefly, the FTC study is a pretest-posttest design in which subjects could not be randomly assigned to experimental and control groups. In the language of Campbell and Stanley (1963), it is a special case of a quasi-experimental design known as the nonequivalent control-group design. In the true experiment with random assignment and effective control, the only rival explanation to the treatment being the cause of the results is sampling error. However, in the nonequivalent control-group design, both sampling error and possible selection bias are plausible rival hypotheses to account for the treatment effects obtained.

The FTC analysis and the Stroud reanalysis both relied heavily on the ability of models in the general tradition of analysis-of-covariance (ANCOVA) to rule out rival hypotheses by making linear adjustments in treatment and control group means for the effects of covarying factors. However, if growth is occurring in the dependent variable (that is, in the verbal and mathematical abilities measured by the SAT), then one of the key ways in which nonrandomized treatment and control groups might differ prior to any coaching intervention is in the rate of this intellectual growth. If the treatment and control groups do differ in growth rates prior to coaching and if these differential rates can be estimated from time series data, even if only linearly from two testing occasions prior to coaching, then another kind of adjustment model can be employed to correct for those self-selection effects embodied in differential group growth that were not predicted from the available covariates used (Bryk & Weisberg, 1976, 1977). Using such a growth model, the reanalysis presented in this section shows in some detail that the Verbal data from the FTC study indeed display differential group growth rates and that estimates of the coaching treatment effect are subject to growth adjustments beyond those predicted from available ANCOVA covariates. The Math data, on the other hand, do not exhibit differential group growth and since the growth model does not adjust for group differences on background variables unrelated to differential growth, the Math coaching effect is better analyzed by ANCOVA.

More specifically, the FTC analysts did not investigate the possibility of differential group growth rates for coached and uncoached students as a rival hypothesis for part of the score increases observed. Campbell (1969) suggests that if two groups start out at time 1 with divergent means, those with the higher mean very likely mature or grow at a greater rate than those with the lower mean. Campbell calls this the interaction of selection and maturation. The FTC analysis of the group demographic variables provides evidence that coached students represent a self-selected population different from uncoached students in a number of ways, including having higher mean pretest scores. The different groups are members of different populations living in different environments. The different environments interact with differences in ability to create and maintain different levels of performance and different rates of growth, suggesting that the use of a statistical model incorporating differential group growth rates is called for.

An interrupted time series design comprising three data points was used to estimate the relative growth rates of coached and uncoached groups during a period prior to the formal coaching intervention. That is, the following design used observations (O) of test performance at time t_0 and time t_1 to estimate growth rate in the absence of a formal intervention (I) for both the treatment group (P) and control group (C). Fortunately, the majority of individuals who attended the most successful coaching school (School A) had three testing sessions, with a coaching intervention taking place between time t_1 and t_2.

Coaching Program Group	(P)	O_0	O_1	I	O_2
Control	(C)	O_0	O_1		O_2
		t_0	t_1		t_2

Figure V-1 presents means and standard deviations of the Verbal test scores at three points in time. Similar data are presented for Math test scores in Figure V-2. The three testings consisted of the PSAT, the SAT taken for the first time in April as a junior, and the SAT taken for the second time in November as a senior. Only one of the three coaching schools in the FTC data set had a large enough student population to furnish sufficient subjects with all three data points. The coached sample included 192 individuals while the control sample numbered 684.

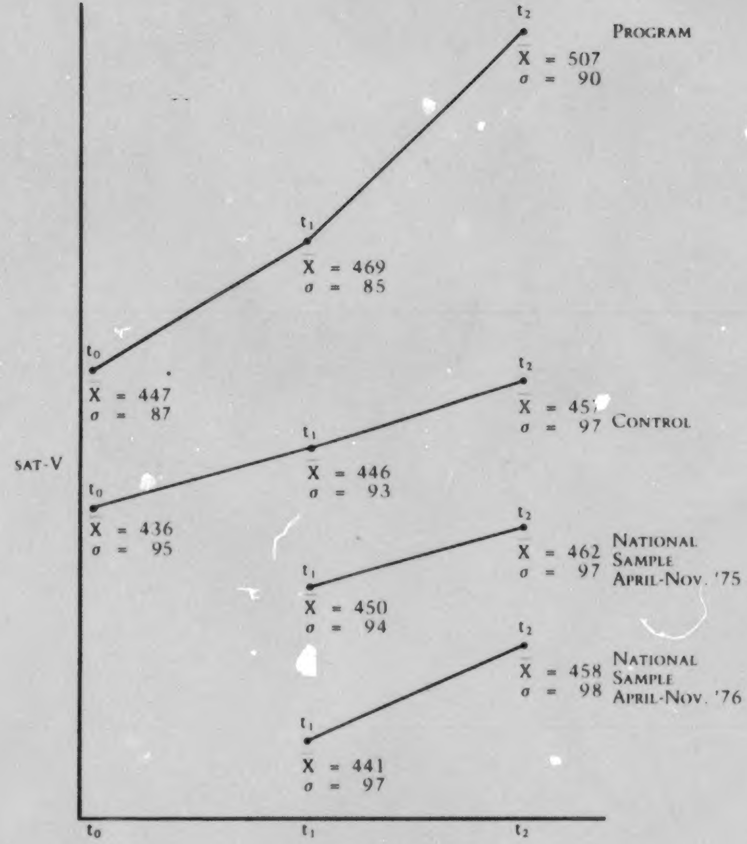


Figure V-1

Verbal Gains Over Three Time Periods for Four Populations

The pattern of verbal data in Figure V-1 suggests that the observed group differences at posttest may well be reflecting differential growth rates as well as treatment effects. Not only is the classic growth phenomenon of increasing means demonstrated in Figure V-1 for both the coaching program students and the controls, but it can be seen that the slope of the line from the PSAT administration (t_0) to the first SAT administration (t_1) is considerably steeper for the coached group than for the control group. The difference in mean score from t_0 to t_1 for the coached group is 22 points while the difference for the uncoached control group is 10

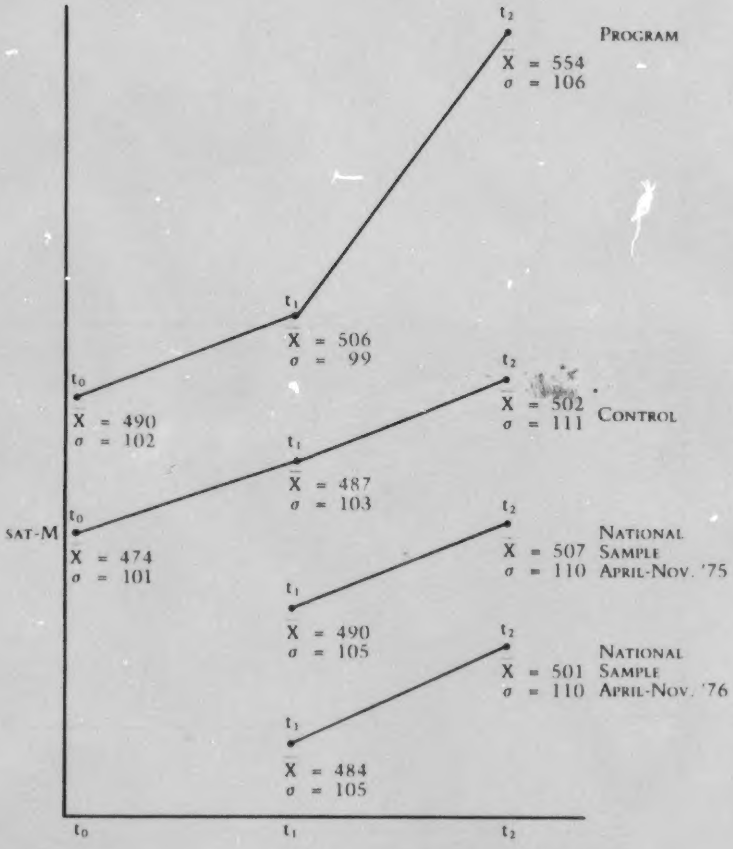


Figure V-2

Math Gains Over Three Time Periods for Four Populations

points—a discrepancy of 12 points. On the other hand, while there appear to be significant differential group growth rates in the Verbal area, this does not seem to be the case for Math (see Figure V-2). This is not unreasonable since mathematics is a skill which requires practice if students intend to maintain or increase their level of achievement. In SAT-Math, the slopes of the lines for coached and control groups are nearly parallel, with the mean difference in Math score from t_0 to t_1 for the coached group only 3 points larger than that for the control group.

In addition, mean scores for the national College Board SAT sample tested in April of the junior year (t_1) and retested in November (t_2) of the senior year are also plotted in Figures V-1 and V-2. These plots represent the 1975 and 1976 populations. Inspection of the verbal plots (Figure V-1) suggests that gains of 12 to 17 points from April to November appear to be commonplace for junior to senior retesters. The fact that the control sample shows a gain of 11 points over the same interval from t_1 to t_2 suggests that they are reasonably representative of junior to senior retesters. Similarly, in Figure V-2, the April to November gain in SAT-Math for the control group is consistent with the 17-point gain for the 1975 and 1976 junior to senior retesters in the national sample. The control group's rate of gain is consistent and linear over both plotted time periods t_0 to t_1 and t_1 to t_2.

An adjustment index may be derived from these data which takes into consideration differential group growth rates in the absence of formal intervention. If the coached group grows at a faster rate than the control group prior to intervention, as might be expected by virtue of its higher initial mean score, the adjustment index (b^*) in the growth model is greater than unity, in contrast to the adjustment index (b) in the traditional ANCOVA model for test-retest data which is typically somewhat less than unity. Using adjustment indices derived from the FTC data, treatment effects were estimated for the growth model and contrasted with the effects obtained from the standard ANCOVA model.

Table V-1 presents ANCOVA and growth-model adjustment indices and coaching effects, as well as the mean growth rates for the coached and control students for Verbal and Math. These estimates of growth are based on the time period between the first and second testing (t_0 to t_1) and thus are not confounded with the coaching intervention. The estimates of growth rates assume linearity of growth, which would seem to be a reasonable assumption given the restricted time period and, even more importantly, given that the observed data for the control group on both Verbal and Math conform quite closely to a linear growth model. If one were to expect a deviation from linearity, it would more likely be in the direction of a steeper slope (in the absence of intervention) between the second and third testing since the motivational level would be at least as great during this time period as in the previous time period between taking the PSAT and the junior-year SAT.

The estimates of the coaching effect under the growth model are about 17 points on Verbal and 31 points on Math. This finding that the Verbal effect is substantially less than the Math effect is in marked contrast to the ANCOVA estimates of quite comparable V and M effects. The growth model appears to yield an estimate of the Verbal coaching effect that is more consistent with earlier studies and with expectations that Math, being generally more curriculum related than Verbal, might be more responsive to coaching or special preparation.

The results for SAT-Verbal in Table V-1 call into question the adequacy of the standard ANCOVA to correct for selection effects when they are present in the form of differential group growth rates not predicted from available covariates. The ANCOVA adjustment index (b) is approximately 1.0 for the Verbal data and somewhat paradoxically 1.25 for the Math data, which is the one situation displaying little or no evidence of differential group growth. However, since the growth rates appear to be static for the Math data, the ANCOVA model is probably the more defensible approach for estimating the coaching effect for Math because ANCOVA controls for all available measures of pre-existing difference, not just those related to differential growth.

In summary, then, examination of testings at three points in time suggests that: (1) the traditional ANCOVA approach used by the FTC is inadequate for the Verbal data because of self-selection effects which are at least partially captured in differential group growth rates; (2) a more appropriate growth-related adjustment

Table V-1

Mean Growth Rates, Adjustment Indices, and Estimated Coaching Effects Under Different Model Assumptions

	Average Growth Rate[1] *(Points/Month)*	*Adjustment Indices and Their Estimates of Coaching Effects*[1]			
		Growth Model		ANCOVA	
	μ_*	b_i^*	a_i^*	b_i	a_i
Verbal					
Program	3.294	1.485	16.946	.993	28.065
Control	1.654				
Mathematics					
Program	2.354	1.102	30.558	1.252	27.625
Control	2.048				

[1] The average gain in points per month is estimated in the absence of intervention. Growth model estimates b_i^* and a_i^* are the adjustment index and the estimated effect based on the group growth rate estimates in the first column. The ANCOVA estimates are the standard estimates arrived at using the list of control variables presented in Table 2, Appendix 3. The ANCOVA b_i is the net adjustment index, when used in the usual ANCOVA equation it yields the effect estimate a_i.

model yields Verbal coaching effects about one-half the size reported by the FTC; and, (3) the Math data are more consistent with the ANCOVA model. Thus, the resulting FTC estimate of the coaching effect in Math is more likely to be reasonable given the available control variables than is the growth model estimate, since the latter does not adjust for group differences in background variables unrelated to growth. This is not to say that the FTC estimates of the Math coaching effect, as well as those of the present analyses, are not overestimates (or underestimates), since the only self-selection causes that have been adjusted for were those reflected in differential growth rates and/or available demographics.

VI

Implications for Educational and Testing Policy and Practice

In summary, the FTC study of commercial coaching found negligible effects for students attending one coaching school and combined coaching and self-selection effects of about 20 to 30 points on both SAT-V and -M for students at another school. The reanalysis by T.W.F. Stroud, using a more sophisticated analysis-of-covariance design and all three coaching schools in the FTC data set, yielded similar overall results: combined coaching/self-selection effects in the neighborhood of 20 to 35 points for both SAT-V and -M at one school and inconsistent and negligible effects at the other two schools. The effects due to coaching per se at the one apparently effective school are probably lower than this, however, because of the confounding with self-selection factors that influence both attendance at that coaching school and performance on the posttest SAT. As we have seen, for example, when factors related to differential group growth rates in the treatment and control groups were taken into account in the reanalysis conducted by D.A. Rock, the combined coaching/self-selection effect for SAT-V dropped to about 17 points while that for SAT-M, which did not exhibit differential group growth rates in these data, remained at about 30 points. But it is impossible to determine with any confidence whether the effects obtained in the FTC study may be attributable in whole or part to uncontrolled self-selection factors rather than to any impact of the coaching program as such.

Thus, overall, the FTC study appears to reveal considerable variability in the coaching/self-selection effects associated with coaching-school attendance, with an estimated combined effect for students at one school of about 20 to 30 points on SAT-Math and very likely about half to two-thirds that (as reflected in corrections for differential group growth rates) on SAT-Verbal. In addition, the sporadic emergence of significant interactions indicates that particular types of students, such as those highly motivated to achieve, or students with particular cultural backgrounds, such as blacks or Asian Americans, might sometimes exhibit larger score increases in some commercial coaching programs. Further research is needed on this problem since the samples of blacks and other minorities in the FTC study were exceedingly small and somewhat atypical. On balance, however, the overall findings for commercial coaching appear generally consistent with the results of prior studies on the effectiveness of special preparation programs offered by high schools, especially the longer and more intensive ones.

To pursue this latter conclusion in more detail, let us inquire how the results of the FTC reanalyses jibe with the rankings of prior coaching studies in regard to student contact time and magnitude of score effects, as summarized in Table II-3. The coaching program at School A in the FTC study entailed 40 hours of student contact time while that at School B involved 24 hours. If it is assumed that roughly half that time in each case was devoted to Verbal coaching and half to Math coaching, then the Verbal student contact time of School A would receive a rank of 8 when added to the rank-order correlations of Table II-3, while the associated Verbal combined coaching/self-selection effect (using the weighted average of estimates from Table IV-1) would receive a rank of 7. The rank for the Math student contact time at School A would be 5, while that for the associated Math combined coaching/self-selection effect would be 4. The corresponding values for School B would be rank 10 for Verbal student contact time and rank 13 for combined SAT-V effect, and rank 7.5 for Math student contact time and rank 11 for combined SAT-M effect. The consistency in these associated ranks between student contact time and magnitude of combined coaching/self-selection effects, especially for School A where the associated ranks differed by only one ordinal position, suggests that the FTC results are of a comparable order with the results of prior studies. With the values for School A included, the new correlations between rank-order of student contact time and rank-order of score effect are .77 for SAT-V across 18 studies and .78 for SAT-M across 11 studies. Further, if the statistically unreliable values of School B are added, the new correlations become .77 for SAT-V across 19 studies and .74 for SAT-M across 12 studies.

If it were to be shown that relatively intensive coaching could substantially improve scores for some students on the SAT, this would have important implications for both educational and testing practice. In considering what these policy implications might be, it would be important to know whether any increased test scores attributable to coaching represent stable improvements in the verbal and mathematical reasoning abilities measured by the SAT or whether they reflect improved facility in overcoming inadvertent sources of test difficulty unrelated to these reasoning abilities, such as difficulty associated with test anxiety or unfamiliarity with different item formats or test-taking strategies. The former case would threaten neither the interpretive validity nor the predictive utility of the SAT, for any genuine long-term improvements in the abilities measured should lead to score increases and should also serve the student well in coping with the demands of college learning, which typically entail the same kinds of verbal and quantitative reasoning by analogy, induction, and deduction as are measured by the SAT. The latter case, by virtue of pointing to construct-irrelevant test difficulty, has double-edged implications: If improved test wiseness increases test scores that were inaccurately low because of construct-irrelevant difficulty, a more accurate assessment of ability level would result and the predictive value of the test should be enhanced. On the other hand, if improved test wiseness leads to test scores that are inaccurately high, the interpretive validity of the test would be diluted and its predictive validity jeopardized. In any particular instance, of course, some combination of improved abilities and improved test wiseness might contribute to improved test performance.

None of the studies of coaching or special preparation conducted so far has systematically addressed this issue of improved abilities; rather, they have all focused on the prior issue of first identifying any score increases associated with coaching programs before inquiring into their causes.* This second step has never been taken and, as a consequence, at this point we can only conjecture as to the likelihood that obtained score increases reflect improved reasoning skills. On the one hand, the verbal and mathematical reasoning abilities measured by the SAT develop over many years of experience and use in nonschool and school settings, and it is difficult to improve them markedly with short-term interventions in late adolescence or young adulthood. On the other hand, the more productive special preparation programs were relatively intensive—the one commercial coaching school where students exhibited significant average effects, for example, involved 10 four-hour sessions plus homework, which is about the equivalent of a one-semester college course. And the effect on scores, while valued, was not marked. Whether due to coaching or other factors, 20 score points or so on a score scale ranging from 200 to 800 with a standard deviation of 100 amounts to fewer than three additional items correct. It is thus quite consistent with the notion of gradual development of reasoning abilities to argue that intensive study over many sessions might yield modest but genuine improvements in reasoning skills.

---

*Although no coaching study has directly addressed this basic question of whether score increases reflect improved abilities or increased test wiseness, Marron (1965) did broach the issue indirectly by investigating the extent to which SAT scores after special instruction predicted freshman class standings at the U.S. service academies and at selective colleges. This appraisal is indirect because it requires not only that the verbal and mathematical reasoning abilities be improved by instruction but that they be effectively utilized in college performance, whereas there are numerous countervailing factors, athletics and extracurricular activities among them, that contribute to scholastic underachievement. Thus, high correlations would provide positive evidence that increased test scores reflect improved abilities being well utilized in college performance, but low correlations would constitute negative evidence only if plausible rival explanations for poor academic performance could be discounted. As a further complication, relatively crude and approximate methods of equating were required in Marron's study to obtain some semblance of distributional comparability across the service academies and across the widely diverse colleges that the preparatory students dispersed to, thereby rendering the outcome admittedly tenuous. Nonetheless, the findings suggest that at the service academies the students did less well academically than the test scores predicted, whereas at the selective colleges the distributions of class standings and test scores did "not seem to be inconsistent" (p. 22).

It would be important to investigate whether, and to what extent, coaching or intensive special preparation programs may lead to improved abilities and, in any event, to consider how secondary schools might be able to improve their students' academic abilities over the longer haul. Since reasoning abilities facilitate educational accomplishments, school programs stressing ability development as well as subject-matter learning should have a synergistic effect on current and subsequent educational achievement. The verbal and mathematical reasoning abilities measured by the SAT are related, after all, not only to college performance but to high school performance as well. The function of the SAT is to measure validly and uniformly across different groups and different settings the current level of developed scholastic abilities facilitative of academic learning, in an effort to improve the accuracy of forecasting subsequent academic performance at the college level. Whether the current level of developed abilities facilitative of college learning derives in part from intensive special preparation programs pointed toward test performance or from long-term general preparation programs pointed toward school attainment or from extensive experiential learning in nonschool settings—that is, from coaching or instruction or experience—is indistinguishable to the SAT. The issue of equity of access to coaching programs that are effective by virtue of ability development, if such could be identified, is thus similar to the issue of equity of access to effective school programs or effective life experiences. Thus, coaching or special preparation programs producing increased test scores by improving the abilities measured would have important implications for educational practice and social policy.

In contrast, if coaching or special preparation programs may lead to substantially increased test scores without improving the abilities measured, there are important implications for testing practice. Such an outcome would imply that the test or the testing experience entails unintended sources of difficulty, such as anxiety over being evaluated or unfamiliarity with different item formats or test-taking strategies, that can be overcome by special preparation. Issues of equity of access to such special preparation become important to the extent that individual differences in test-taking skills per se influence test scores.

Although one might expect the typical coaching program to reduce unfamiliarity by drilling the student in different approaches to different item formats and to allay anxiety by providing feedback on effective item performance, the fragmentary research evidence suggests that if this is all that is done the coaching will have little impact on test scores. The coaching or special preparation programs that emphasize drill-and-practice appear to be associated with small effects, if any, whereas programs that include skill-development components tend to be associated with larger effects (which, by and large, seem to occur more for Math than for Verbal). Nonetheless, following accepted principles of good test practice, test makers should strive to reduce construct-irrelevant test difficulty wherever possible—for example, by avoiding arcane item types, complicated instructions, esoteric or culture-specific content, undue speededness, and the like. Substantial efforts should also be made routinely to increase test familiarization and test wiseness, such as the College Board has done by distributing a sample SAT and answer key to all candidates along with strategic advice in the booklet *Taking the SAT*. With well-made tests the concern is not that students will learn how to capitalize on extraneous clues to inflate their test scores, for such clues should be rare and are further reducible through test analysis, but rather that students learn to cope most advantageously with the test as a standardized vehicle for demonstrating their abilities. To ensure the fairness of the admissions testing process, then, test makers should attempt to make clear to all prospective test takers what the test measures, what kinds of items and formats will be encountered, and what kind and level of preparation the test presupposes, as well as providing practice materials and advice on effective test taking. This should be done not just to minimize the effects on test scores of student differences in this regard, but to increase the validity of the test and to enhance the value of the testing experience for each individual.

Finally, what advice does all this imply for the prospective SAT taker? If this advice were based mainly on a summary of the research findings as best we can interpret them at this point, the student might be told something like this:

> The verbal and mathematical reasoning abilities measured by the SAT develop over years of experience and use both within and outside of school. In preparing for the test, a review of mathematical concepts may be beneficial, especially for students not currently enrolled in mathematics courses.
>
> Studies of the effects of drill-and-practice on sample items or of last-minute cramming show little score increase, although such practice may serve to familiarize the student with different item formats and to reduce apprehension about what to expect. The College Board distributes test-familiarization materials to all candidates and recommends that they take the sample SAT provided, carefully review the answers, and read the booklet on *Taking the SAT*.
>
> Studies of students attending special preparation programs available in many high schools show a range of effects averaging less than 10 score points on Verbal and less than 15 points on Math for programs requiring an average of about 20 hours or so of total student contact time. Part or all of these score increases may be due to factors other than the special preparation program offered, such as the level of motivation or prior educational experience of the students. And individual students or certain groups or types of students might experience larger increases or possibly even losses.
>
> A recent study of students enrolled in commercial coaching courses shows variability in average score improvement depending on which course the students attended: negligible effects were found for students attending two of the three coaching schools studied, while average effects in the 20- to 30-point range for Math and perhaps half to two-thirds that for Verbal were associated with attendance at the third school, which entailed about 40 hours of total student contact time. Even for students attending this latter school, however, it was not at all clear what proportion of the score increases was due to coaching experiences as such rather than to motivational and other background characteristics typical of the students taking that course. Again, individual students or certain groups or types of students might experience larger increases or possibly even losses.
>
> There have been no systematic studies comparing the relative effectiveness of different types of coaching programs, whether they be commercial or school-based, let alone their relative effectiveness

> for particular groups or types of students, but the larger score effects appear generally to be associated with longer-term intensive programs emphasizing skill development in verbal and mathematical reasoning.
>
> In any event, whether due to coaching or to personal factors, score increases of 20 to 30 points on a 200- to 800-point scale correspond to approximately three additional items correct, and comparable or larger effects might be obtained if students were to devote the same level of effort on a regular basis to classroom instruction and independent study.

Although further research is needed to clarify the nature of the relationship, it seems very likely that improvement of the comprehension and reasoning skills measured by the SAT, when it occurs, is a function of the time and effort expended. Furthermore, each additional increase in ability may require increasing amounts of time and effort, probably geometrically increasing amounts. But since the time required to achieve average score increases greater than 20 to 30 points or so, especially in Verbal, may rapidly approach that of full-time schooling, the soundest long-range mode of preparation for the SAT would appear to be a secondary-school program emphasizing the development of thought as well as knowledge.

VII

References

Alderman, D. L., & Powers, D. E. *The effects of special preparation on SAT-verbal scores* (CB RDR 78-79, No. 4 and RR 79-1). Princeton, NJ: Educational Testing Service, 1979. (*American Educational Research Journal*, 1980, **17**, 239-251.)

Belson, W. A. A technique for studying the effects of a television broadcast. *Applied Statistics*, 1956, **5**, 195-202.

Breland, H. M. *Population validity and college entrance measures* (Research Monograph No. 8). New York: The College Board, 1979.

Bryk, A. S., & Weisberg, H. I. Value-added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics*, 1976, **1**, 127-155.

Bryk, A. S., & Weisberg, H. I. Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin*, 1977, **84**, 950-962.

Campbell, D. T. Reforms as experiments. *American Psychologist*, 1969, **24**, 409-429.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally, 1963.

Cochran, W. G. The use of covariance in observational studies. *Applied Statistics*, 1968, **17**, 270-275.

Coffman, W. E., & Parry, M. E. Effects of an accelerated reading course on SAT-V scores. *Personnel and Guidance Journal*, 1967, **46**, 292-296.

Dear, R. E. *The effect of a program of intensive coaching on SAT scores* (ETS RB 58-5). Princeton, NJ: Educational Testing Service, 1958. (Reported in French, J. W. & Dear, R. E. Effect of coaching on an aptitude test. *Educational and Psychological Measurement*, 1959, **19**, 319-330.)

Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.

Donlon, T. F. The effects of special preparation for tests: The ETS experience. Paper in preparation.

Dyer, H. S. Does coaching help? *College Board Review*, 1953, **19**, 331-335. (Reported in, French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. *Educational and Psychological Measurement*, 1959, **19**, 319-330.)

Evans, F. R., & Pike, L. W. The effects of instruction for three mathematics item formats. *Journal of Educational Measurement*, 1973, **10**, 257-272.

Federal Trade Commission, Boston Regional Office. *Staff memorandum of the Boston Regional Office of the Federal Trade Commission: The effects of coaching on standardized admission examinations*. Boston, Mass.: Federal Trade Commission, Boston Regional Office, September 1978.

Federal Trade Commission, Bureau of Consumer Protection. *Effects of coaching on standardized admission examinations: Revised statistical analyses of data gathered by Boston Regional Office of the Federal Trade Commission*. Washington, DC: Federal Trade Commission, Bureau of Consumer Protection, March 1979.

Frankel, E. Effects of growth, practice, and coaching on Scholastic Aptitude Test scores. *Personnel and Guidance Journal*, 1960, **38**, 713-719.

French, J. W. *The coachability of the SAT in public schools* (RB 55-26). Princeton, NJ: Educational Testing Service, 1955.

French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. *Educational and Psychological Measurement*, 1959, **19**, 319-330.

Lass, A. H. Unpublished study. Brooklyn, NY: Abraham Lincoln High School, 1958.

Marron, J. E. *Preparatory school test preparation: Special test preparation, its effect on College Board scores and the relationship of affected scores to subsequent college performance*. West Point, NY: Research Division, Office of the Director of Admissions and Registrar, United States Military Academy, 1965.

Pallone, N. J. Effects of short-term and long-term developmental reading courses upon S.A.T. verbal scores. *Personnel and Guidance Journal*, 1961, **39**, 654-657.

Pike, L. W. *Short-term instruction, testwiseness, and the Scholastic Aptitude Test: A literature review with research recommendations* (CB RDR 77-78, No. 2 and ETS RB 78-2). Princeton, NJ: Educational Testing Service, 1978.

Pike, L. W. *Implicit guessing strategies of GRE-aptitude examinees classified by ethnic group and sex* (GRE Board Professional Report GREB No. 75-10P). Princeton, NJ: Educational Testing Service, 1980.

Pike, L. W., & Evans, F. R. *The effects of special instruction for three kinds of mathematics aptitude items* (CB RDR 71-72, No. 7 and ETS RB 72-19.) Princeton, NJ: Educational Testing Service, 1972.

Powers, D. E., & Alderman, D. L. *The use, acceptance, and impact of Taking the SAT—A test familiarization booklet* (CB RDR 78-79, No. 6 and ETS RR 79-3). Princeton, NJ: Educational Testing Service, 1979.

Roberts, S. O., & Oppenheim, D. B. *The effect of special instruction upon test performance of high school students in Tennessee* (CB RDR 66-7, No. 1 and ETS RB 66-36). Princeton, NJ: Educational Testing Service, 1966.

Rock, D. A., & Werts, C. E. *Construct validity of the SAT across populations—An empirical confirmatory study* (CB RDR 78-79, No. 5 and RR 79-2). Princeton, NJ: Educational Testing Service, 1979.

Slack, W. V., & Porter, D. The Scholastic Aptitude Test: A critical appraisal. *Harvard Educational Review*, 1980, **50,** 154-175.

Stern, J. Personal communication, 1975.

Vernon, P. E. The determinants of reading comprehension. *Educational and Psychological Measurement*, 1962, **22,** 269-286.

Whitla, D. K. Effect of tutoring on Scholastic Aptitude Test scores. *Personnel and Guidance Journal*, 1962, **41,** 32-37.

VIII

Appendices

Appendix 1

Critical Notes on the FTC Coaching Study

Critical Notes on the FTC Coaching Study

DONALD L. ALDERMAN, November 16, 1979

Staff Memorandum of the Boston Regional Office (BRO)

The disclaimers and reservations attached to the Boston memorandum make it quite clear that the Federal Trade Commission (FTC) itself found serious flaws in this staff report. Yet it may be useful to enumerate those criticisms as well as to offer further comments on specific details since the "revised statistical analyses" conducted by the FTC's Bureau of Consumer Protection (BCP) in Washington accepts the data set and repeats several assumptions of the Boston memorandum.

The note on the cover of the Boston memorandum goes further than usual disclaimers in stating that "the Commission specifically believes that some of the conclusions in the study are not supported by the evidence obtained in the investigation." This strong disavowal of certain conclusions may be attributed to "several major flaws in the data analysis," quoting from the BCP's notice to recipients of the Boston memorandum. The notice cites four specific flaws: (1) comparisons are made "for groups of coached and uncoached students without controlling for differences which may exist in personal and demographic characteristics of the students in the two groups;" (2) failure "to provide tests of statistical significance which are necessary to interpret the results" (i.e., whether differences may simply be due to chance occurrences rather than treatment effects); (3) defects that concern "the method used to present the findings from the data analysis" (e.g., discussion of results in terms of a nonexistent subpopulation of students); and, (4) "all the limitations associated with [a nonexperimental design]."

These points stressed by the FTC focus on the statistical analysis and give inadequate attention to the data editing which necessarily took place prior to the stage of statistical analysis. Moreover, the subsequent BCP report repeats several weak assumptions made in the Boston memorandum.

Assumptions and Data Analysis

These additional comments concern some of the BRO's assumptions and some of the steps taken in constructing the final data base. But the design of the study would itself preclude any definitive conclusions. Numerous alternative explanations exist for any differences in the test performance of the two comparison groups (i.e., "coached" and "uncoached") since students in these groups differ markedly on key demographic characteristics (see Table 1 of the BCP's report, pp. 8-11). And the statistical analyses undertaken in the BRO memorandum exacerbate these initial group differences by ignoring them. Nevertheless, I offer a few additional comments:

- *Scope of inferences.* The BRO memorandum assumes that "valid inferences about the coachability of other examinations can be drawn from the specific results we obtain for the SAT and LSAT." Given the pre-existing differences evident in the comparison groups, there is not even a strong basis for inferences about the SAT and LSAT let alone any other examinations. Indeed, the results themselves show inconsistencies across test administrations, commercial schools, and examinations.

- *Representative sample.* The enumeration of the study's assumptions (BRO memorandum, p. 49) begins with a statement concerning the sample's representativeness of the entire SAT and LSAT candidate populations. Although the control group was a random sample drawn from history files, the "coached" group is obviously very different from the population of candidates (see Table 1 of BCP's report). The release of BRO's technical appendix "SDQ," which gives demographic profiles by comparison group, should confirm the weakness of this assumption.

- *Consistency of treatment effects.* Another assumption was that the "coaching school [effect] is consistent during the study period (p. 49)." Certainly the consistency of treatment effects across schools or test administrations should be an open question subject to investigation rather than conjecture or assumption.

- *Treatment self-selection and control contamination.* The strangest assumption made in the BRO memorandum is that "the effects of enrollee self-selection, if any, and of coaching of presumably uncoached students offset one another (p. 49)." There is no way to estimate the extent of treatment contamination in the control group since there were no data collected on this issue, and the BRO memorandum made no attempt to identify a component of score change which might be attributed to self-selection in the treatment group. Yet the statement appears that these unknown and unspecified effects offset one another.

- *Data editing.* The steps which led from the ETS tapes and the commercial schools' enrollment rosters to a final data base were difficult to follow in the BRO memorandum. If my reconstruction of these steps is accurate, substantial numbers of students enrolled in commercial coaching courses were ommitted from the final data base:

	Number of Students	
	LSAT	SAT
Enrollment in commercial course	11,906	2,286
Matches with ETS test files	9,029	1,777
Students with results from one or two test administrations	6,894	
Students with GPA (LSAT) or PSAT (SAT)	4,662	
Students with known sex and ethnicity	2,830*	
Final "study group" in treatment (coached) condition	3,509*	603

*(p. 58)

Such sizeable losses (and gains!) raise questions about possible changes in the nature of the sample at each step of the data editing. It has already been noted elsewhere that the number of commercial school enrollees without SAT testing histories seems high (N = 509). But the reduction in sample size from the number of cases with testing histories to the final "study group" is twice as large.

Revised Statistical Analyses of the Bureau of Consumer Protection

The BCP report is a much stronger technical document than the BRO memorandum. Its conclusions must be subject to the same qualifications since the study necessarily involves comparisons against a nonequivalent control group, but the BCP report acknowledges these limitations and takes care to qualify any inferences from the results. Furthermore, the explanation of the steps taken in editing and in analyzing the data is far better than that contained in the BRO memorandum. Nevertheless, they indicate that SAT testing histories were identified for 1568 coaching school enrollees but were not located for "approximately 600 individuals identified from the coaching school list" (p. 5). This attrition rate is about 28 percent and represents a significant loss. Moreover, the response rate for the SDQ for the coached and uncoached students (64 and 67 percent, respectively) is considerably lower than that typically found for SAT candidate groups—about 86 percent. Such high attrition rates from each source may have introduced biases relevant to coaching and its effects.

The BCP report, despite its inherent limitations, provides information on a topic of concern and interest. On the whole, we should welcome the report as a contribution to the literature on coaching. The effects (see BCP report, Table 3, p. 20) correspond to three additional items correct on each of the SAT-V and SAT-M and probably incorporate several components, including self-selection, instruction, and coaching. Moreover, those effects held just for School A. School B had negligible impact on SAT scores. If we view or demonstrate these effects as probable overestimates arising from comparing nonequivalent groups and confounding several possible sources of score changes, three additional items correct seems not to represent a dramatic gain. The same effect might be found for students devoting comparable effort and time to independent study or regular academic courses covering typical subject matter.

Comments on selected points in the BCP report follow:

> *"Underachievers."* The BCP report suggests that coaching at School A was "effective" for "underachievers:" an estimated treatment effect of 25 points was found at School A for students whose SAT scores were lower than might be expected on the basis of their high school grades, class rank, and other demographic characteristics. Usually students whose school grades seem high in relation to their test scores would be called "overachievers." And that term, "overachievers," seems more consistent with the demographic profiles in Table 1 (BCP report, pp. 8-11); coached students had higher class ranks (31% in the top tenth of their class), reported larger parental incomes (41% with $30,000 or more), attended nonpublic schools more often (45% in nonpublic schools), had higher English grades (55% with an A in their latest course), and had higher mathematics grades (48% with an A in their latest course) than control students. These students in the coached group were not underachievers or disadvantaged, they were overachievers and middle-class.

Adjustment for self-selection. The dropping of PSAT as a covariate (and the concomitant increase in the standard error of the estimated treatment effects) in attempting to compensate for self-selection left me puzzled. After confirming students' self-selection into the coached group by noting unaccounted differences in PSAT (BCP report, Table 4, p. 26), PSAT is dropped as a covariate in the regression analyses. If we extend that argument to all other independent variables which show significant differences between the coached and uncoached groups (BCP report, Table 1, pp. 8-11), the only covariates left would be sex and expected years of English. The point is that all of the variables, dependent and independent, show self-selection and no statistical analysis will fully compensate for the nonequivalent control group.

True experimental design. The BCP report takes the position that a true experiment with random assignment of students to comparison groups would be unethical ("as it would require denying access to commercial coaching to students who want it" [BCP report, p. 3]), expensive, and time-consuming. But our own most recent study of special preparation for the SAT-V did follow a true experimental design by delaying rather than denying access to the treatment.

The FTC memorandum from the BRO and report from the BCP do not close the question of the impact of commercial coaching enterprises. It has not been unequivocally shown that such programs cause score gains on national admissions tests. Indeed, no amount of research can disprove the possible effectiveness of coaching. The possibility will remain open since there is always another program or a different student population. Even an experimental study showing no significant effects this year would not be conclusive evidence regarding new programs and different students next year. But we can perhaps clarify this issue by explaining sources of score changes on major admissions tests and by studying the impact of regular academic programs.

Critical Notes on the FTC Coaching Study

Donald E. Powers, November 19, 1979

The claims made for the efficacy of special preparation or "coaching" by those commercial or proprietary schools that offer such preparation programs for standardized admission tests is a proper concern of the Federal Trade Commission (FTC). The Commission deserves support for addressing this thorny issue and for assembling what appears to be the first credible set of data from which inferences about the effects of commercial coaching reasonably could be attempted. As a result of the FTC effort we now know something more about the effects of commercial coaching. Through further analyses that are possible with the FTC data base, even more may be learned.

The FTC should also be applauded for acknowledging the limitations inherent in the type of quasi-experimental study that was conducted, for recognizing the serious flaws in an earlier report issued by its Boston Regional Office, and for attempting to rectify some of these flaws through reanalyses.

Nevertheless, with respect to the more technically adequate report issued by the FTC's Bureau of Consumer Protection (BCP) in Washington, a number of comments can be made regarding: (1) the general reporting of results; (2) the assumptions underlying various analyses and findings; (3) the design and analysis strategies employed; (4) the testing of hypotheses; (5) the logic involved in drawing inferences; and, (6) the reasonableness of results. Needless to say, perhaps, these categories may have substantial overlap.

Reporting

In terms of style, the BCP report is generally well-written, with a minimum of technical jargon. However, four relatively minor points might be made.

- The report's title, "Effects of Coaching on Standardized Admission Examinations," seems to suggest that the results of the study pertain to a variety of standardized admissions tests, which include both aptitude and achievement measures, and to a variety of coaching methods and courses. Since the study has been restricted to specific types of coaching and to one particular aptitude test (the SAT), a title more adequately reflecting the narrower focus of the study would have been more appropriate.
- The report's executive summary is selective in focusing on the results at only one of the two schools involved. Inclusion of a statement concerning the lack of effectiveness of the second school would have been appropriate.
- The report's summary of an average effect of 25 points "on both the verbal and math SAT exams" at one school obscures some important information about the effects, namely that the effect on Verbal scores was somewhat higher than the effect on Math scores.
- Detail for some especially relevant descriptive information is omitted from the report. Table 1 of the BCP report, which shows demographic profiles of coached and uncoached students, fails to provide the statistics for PSAT scores, which are acknowledged in the narrative to differ for coached and noncoached students. The omission of data on the PSAT is glaring because of the report's contention that self-selection is based primarily on these earlier test scores.

Assumptions

Like most quasi-experiments, the FTC study rests on a number of assumptions, both explicit and implicit, that can be examined for reasonableness:

- The BCP report states on page 2 that:

> Conversely, a non-experimental control group may lead to an un-

> derestimate of coaching benefits, that is, the uncoached students may in fact have received some form of coaching other than formal enrollment in a commercial coaching course. They may, for example, have attended a course offered by a school in the not-for-profit segment of the industry or have engaged in extensive self-preparation. These unmonitored efforts, if they occur and if they are effective, would tend to increase the average test scores of the "uncoached" students. These increased scores, containing a component properly attributable to coaching, would tend to shrink the apparent benefit from commercial coaching.

In the absence of any data, this assumption would seem reasonable. However, some reworked data (Table 1) from the study of the effects of the booklet *Taking the* SAT suggest that coached students may be more likely than uncoached students, except for the use of College Board publications, to also engage in other methods of preparation. Thus, by FTC reasoning, their analyses would actually be more likely to *overestimate* the effects of coaching by failing to account for the other ways in which coached students may have prepared.

Moreover, the fact that coached students may prepare more diligently than uncoached students would seem to suggest a motivation to do well on the SAT that was probably not captured by any of the variables included in the FTC study. (It should be noted that the coached students in Table 1 reported "attending coaching sessions outside school," but not necessarily *commercial* coaching schools of the type studied by the FTC.) Hence, the reasonableness of this FTC assumption that preparation efforts on the part of the uncoached students tended to shrink the apparent benefit from commercial coaching may be questioned.

- The assumption that, compared to nonexperimental studies, experimental studies of the effects of special preparation are impractical, unethical, expensive, and time-consuming can be questioned by citing the recently completed experimental study of special preparation programs offered by secondary schools by Alderman and Powers.

- The FTC study attends to the importance of self-selection, but apparently implicitly assumes that all students use the same basis for choosing to attend coaching schools, i.e., that self-selection is uniform. This assumption attributes a simplicity to self-selection that may not be warranted.

Table 1

Use of Various Preparation Methods by Coached and Uncoached Students

	Method of Preparation																	
	(1) *Answered Questions in About the* SAT		(2) *Read Taking the* SAT		(3) *Reviewed Math on Own*		(4) *Reviewed Vocabulary or English on Own*		(5) *Reviewed Other Material on Own*		(6) *Read Test Preparatory Books*		(7) *Attended Preparatory Session at School*		(8) *Attended Coaching Outside School*		(9) *Other*	
Group	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Coached N = 62	46	74.2	28[1]	82.3	34	54.8	43	69.4	10	16.1	44	71.0	17	27.4	62	100.0	10	16.1
Uncoached N = 1279	978	76.5	563[2]	92.0	361	28.2	545	42.6	110	8.6	636	49.7	189	14.8	00	0.0	147	11.5
Total	1024	76.3	591[3]	91.5	395	29.5	588	43.8	120	8.9	680	50.7	206	15.4	62	4.6	157	11.7
χ^2	1.69		3.84		20.16		17.18		4.10		10.67		7.27		—		1.23	
p	p = n.s.		$p < .05$		$p < .001$		$p < .001$		$p < .05$		$p < .05$		$p < .01$				p = n.s.	

[1] Base is 34. [2] Base is 612. [3] Base is 646.

Source: Powers, D. E. *Students' Use of and Reactions to Alternative Methods of Preparing for the* SAT. In progress.

Design and Analysis

The FTC presents the inherent limitations of the non-experimental nature of the design used. The application of multiple regression analysis as the primary data analysis technique seems appropriate. However, the following additional refinements in analysis might have been useful:

- A more careful entry, perhaps serially, of variables or sets of variables in the analyses, instead of entering all of them simultaneously, might have been even more informative. A careful examination of the collinearity of variables might also have been desirable.
- Additional analyses of possible interactive effects might have proven useful. The FTC researchers did test for PSAT coaching interactions and concluded that they need not be taken into account. However, a number of other interaction effects could have been included in the analyses, such as the interaction of coaching with:

  —number of previous PSATS

  —number of years of math, and

  —high school grades.
- Including more interaction terms would presumably have resulted in a better feel for the joint effects of coaching and other important variables or the conditions under which coaching is most effective.
- Separate analyses for each coaching school, and/or more sophisticated procedures for pooling data, could have been conducted, given the acknowledged differences between the students attending the two coaching schools. Important effects may have been obscured by merely combining together data for all students at both schools as in the demographic profiles.

Hypothesis Testing

At times the FTC study is too selective in the testing of plausible hypotheses. This defect is especially evident in the analyses of self-selection bias.

- As one example, on page 18 of the BCP report, the hypothesis is presented that students who are coached before they take the

SAT for the first time are serious and do not plan to take it a second time, while those who are not coached are not as serious because they do plan to take it a second time. This hypothesis is tested by showing that coached students were *more* likely than their uncoached counterparts to take the SAT a second time and that, therefore, the issue was not a cause for concern. However, no consideration is given to the counter-hypothesis that because coached students were more likely than uncoached students to take the exam a second time, they are *more* serious and *more* determined to persist until they obtain satisfactory scores.

- The FTC study acknowledges the potential for biased estimates of effects resulting from student self-selection. However, the study focuses on only one possible type of self-selection bias, i.e., that students choose to attend coaching schools on the basis of how their first score (SAT or PSAT) compares with what might be expected from their high school rank, their parents' income, etc. This hypothesis is tested by regressing first SAT score (and PSAT score) on demographic variables and showing that coached students, on the average, obtained slightly lower than expected SAT and PSAT scores, whereas noncoached students did slightly better than expected. However, similar analyses using each of the demographic characteristics (instead of PSAT and SAT scores) as dependent variables might very well have shown similar, or even more dramatic, differences between coached and uncoached students, suggesting self-selection on other variables.

One plausible alternative hypothesis is that self-selection operates primarily on the basis of family income, an hypothesis that is supported by the demographic profiles which show a very strong relationship between attending coaching schools and family income. It seems highly likely that regressing family income on other demographic characteristics and PSAT scores would show that coached students come from families whose incomes are higher than expected from other demographic characteristics, and that, therefore, self-selection might well be based primarily on parental income (i.e., on one's family's ability and willingness to pay for coaching). The same analyses could be applied to any of the other available student characteristics.

The following observations suggest that self-selection on the basis of family income is not only consistent with the FTC data but also at least equally plausible to self-selection based on expectations for PSAT (SAT) scores:

(1) One wonders how precisely students can estimate, or develop strong expectations for, how well they should perform on the PSAT or SAT according to their personal demographic profiles, i.e., how closely statistical expectations correspond to student expectations. As an example, would a student having a particular demographic profile (white male ranking in the top tenth of his class, from a family earning $21,000 a year, having taken four years of math and obtaining an excellent grade in his latest math course) expect to get a PSAT-M score of 56, as would be predicted from the FTC-generated regression equations? How likely would he be to seek coaching if his actual score was below this, or to forego coaching if his actual score exceeded 56?

(2) Differences between coached and noncoached students on variables other than family income could be explained on the basis of incidental selection. If students were to base their decisions to attend coaching schools solely (explicitly) on their parents' ability to pay, differences on most other demographic variables (e.g., PSAT scores, grades, attendance at private schools) could be merely incidental because of the positive correlation of each of these variables with parental income.

(3) Differences in self-selection were detected for students attending the two coaching schools in the FTC study. Self-selection of the type hypothesized in the study was noted for School A, which charges $225, but not for School B, which charges only $75. This pattern is consistent with the alternative hypothesis that family income is the major self-selection variable.

Testing other types of self-selection bias seems very important in light of the FTC conclusion that coaching is effective for "underachievers," which is defined in the report as being the same as self-selection.

Logic and Inferences

Several of the conclusions drawn in the FTC report are open to further consideration.

- The analysis of self-selection bias shows that coached students had unexpectedly low scores on *both* a previous SAT and the PSAT, from which it is concluded that students who tend to enroll in coaching schools are those who do worse than expected on standardized tests. The fact that, on the average, those choosing coaching schools do worse than expected on *both* previous exams (SAT and PSAT) is used to infer that self-selection is not on the basis of scoring lower than expected by chance (i.e., that because the phenomenon is noted for two tests, it is not by chance).

  However, this analysis rules out only that self-selection is not based solely on obtaining by chance an initial SAT score that was lower than expected. It does not rule out the hypothesis that self-selection is based on scoring low by chance twice (on both the PSAT and the first SAT). That is, those students who are unlucky twice may be the ones choosing to select coaching. Using the FTC technique, we can reason that if, say, about one-fifth of all students tend to be unlucky, i.e., score below expectations on each exam, then about one in twenty-five would be expected to be unlucky both times. Given the relatively small percentage of all examinees who attend commercial coaching schools, coached students could very well come primarily from this doubly unlucky group.

- A major conclusion of the study that rests on the analysis of self-selection is that "coaching can be effective for those who do not score well on standardized tests" (p. 35), and that "if only underachievers can be helped, it is possible that coaching at the second school would be effective for such students" (executive summary). Because of its policy implications, this conclusion deserves critical analysis. The reasoning from which this conclusion results is faulty because self-selection of the test underachievement type was noted for only one of the two schools. That is, there are underachievers at one school but not at the other. This confounding of particular school with type of student does not allow the type of conclusion made in the FTC report. It is not possible to conclude from the FTC analysis of only two schools whether the nature of the program or the type of student in attendance is responsible for the apparent effect on test scores. An analysis of the effects for under- and overachievers *within* schools is needed to make an inference of this nature.

- Related to the major conclusions listed above is the extremely weak argument that because self-selection was not detected at School A for one exam (the 1975 Verbal exam, for which coaching was effective), then coaching at School A can be effective for all students, not just underachievers. Not only is this inference not justified, but the reasonableness of the finding upon which it is based [i.e., that self-selection was not detected for one portion (Verbal) of one exam (the 1975 administration) can be questioned]. It is difficult to understand how or why self-selection should be different in one year for only one part of the exam.

Findings

A number of the report's findings, while possibly correct, can be classified as either counterintuitive, curious, or needing further explanation or clarification.

- The finding that the effects of coaching on the Verbal part of the SAT were as large as (in some instances larger than) the effects on the Math section is counterintuitive to what we know about the development of verbal and math abilities as measured by the SAT. This somewhat unexpected finding could have been given further attention in the report.

- Table 3 on page 20 of the FTC report shows two estimates of effects that are significantly different statistically for subsamples, but these differences are not mentioned in the report. The first is the difference between estimates for Verbal effects for subsamples 1 and 2 for School A, suggesting that School A was more effective in 1976 than in 1975. Since School A was well established before 1975, this difference may deserve some attention.

  The second significant difference is between subsamples 5 and 6 for Math at School A, suggesting that School A's instruction is more effective for second-time SAT takers than for first-time SAT takers. Possibly, these significant differences merely reflect the relatively large number of significance tests. However, at least the very large difference in year-to-year effects on Verbal scores for School A should have received some mention. Or should we believe that such fluctuations are typical of this kind of study?

- As mentioned above, the analysis of self-selection discussed in the final section of the report concludes that self-selection was "in effect for School A for each of the test dates for both of the exams, with the sole exception of the 1975 verbal exam" (p. 35). This is a curious finding, since there should be no reason to believe that that self-selection would be different for this particular exam, especially since self-selection was detected for the Math portion of the same administration.

- The finding that School A is more effective than School B deserves more attention. This result seems reasonable since School A offers a longer, and more expensive, course than does School B. However, another hypothesis can be offered to account for the apparent difference in estimates of the effectiveness of the two schools. Since the tuition for School A is three times that of School B (and assuming selection is related to cost), it seems likely that the uncoached control group is more like the students coached at School B than those coached at School A, an hypothesis that can be easily tested with the FTC data.

  If this is so, then the analysis employed by the FTC would be less likely to yield biased estimates of effects for School B than for School A, since the coached and uncoached students would also be less likely to differ on some other uncontrolled variable(s) related to SAT scores and to attending coaching schools.

- Several of the analyses in the FTC report yielded larger, though less precise, estimates of effects for the variable NEWSPSAT (defined as the number of PSATS taken before the first SAT) than for coaching at either School A or School B. These estimates are summarized in Table 2 but not mentioned in the FTC report. Such large (and in the case of subsample 5, significant) estimates of the effect of taking a second PSAT seem, intuitively, much too high. However, if they are correct, they place the smaller estimates of the effects of coaching in a much different perspective. It seems more probable, however, that the large effects attributed to retaking the PSAT may be overestimates because of some other factor that covaries with both SAT scores and the tendency to repeat the PSAT and also quite possibly, with attending coaching schools. It may also result from small samples. In any event, if we are asked to take the results of the FTC analyses at face value, the appropriate (and much less expensive) recommendation to students preparing for the SAT would be simply to repeat the PSAT.

  Finally, such a large PSAT effect does not seem reasonable because most of the coaching courses place a premium on prac-

Table 2

Effects of PSAT Practice and Coaching
(Summarized from FTC report)

		Variable		
Sample No./Test	*Test*	NEWNPSAT *(1 vs. 2 PSATs)*	*Coach 1*	*Coach 2*
2/First	Effect	43.8	44.5	3.5
Verbal	S.E.	54.9	6.0	6.9
2/First	Effect	23.4	26.5	0.2
Math	S.E.	58.4	6.4	7.3
5/First	Effect	41.7	29.7	-1.8
Verbal	S.E.	14.4	4.4	5.7
5/First	Effect	41.3	19.2	5.4
Math	S.E.	14.8	4.5	5.8
2/First	Effect	65.9	26.8	0.8
Verbal (PSAT deleted)	S.E.	83.9	9.2	10.6
2/First	Effect	80.6	7.5	8.8
Math (PSAT deleted)	S.E.	85.4	9.3	10.7
5/First	Effect	68.2	11.5	-2.3
Verbal (PSAT deleted)	S.E.	22.7	6.9	8.7
5/First	Effect	51.7	0.5	13.4
Math (PSAT deleted)	S.E.	22.3	6.7	8.6

tice, which is distributed over time, and might be expected to be more effective than an additional PSAT.

- Table 5 of the FTC report contains an apparent error. The value of the estimate of the impact on the Math SAT for School A, sample 5, has been recorded as 5.5. However, the corresponding regression analysis in the appendix indicates a value of 0.55 (p. C-10).

The fact that the FTC study has stimulated discussion of an important issue and has raised many additional questions may be one of its major contributions. The critical attention that the report has attracted should help us better understand the effects of coaching on admissions tests.

Critical Notes on the FTC Coaching Study

DONALD A. ROCK, December 17, 1979

The FTC analysis (Boston Memorandum) and reanalysis (Washington Report) of the effect of commercial coaching on admissions test score gains applied standard ANCOVA procedures in an attempt to statistically equate two *different* populations. The FTC report itself documented that the two populations differ on both demographic and ability variables. Conclusions based on such analyses are only valid if: (1) strong assumptions can be made about the appropriateness of the statistical correction used in eliminating the measured pre-existing differences, and (2) all unmeasured pre-existing differences are unrelated to observed differential gains over the time periods in question. Subsequent reanalyses at the Educational Testing Service suggest that (1) and (2) above are likely to be untrue in some cases and only partially true in others. Some of the failures to meet assumptions were empirically demonstrated (Appendix 3) and their consequences with respect to interpretations outlined. Problems with the FTC analysis and reanalysis, some statistical and some logical, are as follows:

(1) The FTC authors report that approximately 600 scores for coaching school enrollees were not found in the SAT files. This, I think, has serious consequences for their conclusion. If these 600 individuals cancelled their scores because they did not think they did well, then one would expect the estimates of coaching effects to be biased upwards.

(2) The ETS reanalysis of test scores at three points in time suggests that the standard ANCOVA approach used by the FTC was

inappropriate for the verbal data because of self-selection effects which were at least partially captured in differential group growth rates. More appropriate growth related adjustment models yielded verbal coaching effects about one-half the size reported by the FTC report.

The mathematics data were found to be more consistent with the standard ANCOVA model and thus more likely to yield reasonable estimates of coaching effects given the *available* control variables. This is not to say that the FTC and the ETS reanalysis estimates of the mathematics coaching effect are not overestimates, since the only self-selection causes that have been controlled for were those reflected in differential growth rates and/or available demographics. One missing piece of information is the reliability of the pretest scores. When individuals self-select to treatments, commonly held psychometric wisdom suggests that the standard ANCOVA model can be expected to underadjust to the extent that the pretests are less than perfectly reliable, thereby yielding an overestimate of the coaching effect.

(3) The FTC concludes that the students . . .

> who attended the most "effective" school tended to be underachievers on standardized exams, i.e., they scored lower on standardized exams than would have been predicted given their personal and demographic characteristics (including such factors as grades in school and class rank). If this underachieving was random rather than systematic, the results showing the benefits of the coaching received at School A might have been overstated. Analysis was conducted, however, showing that the underachievement by the students was not due to chance, and probably would have continued in the absence of coaching.
>
> (Executive Summary)

The above conclusions were based on examining the size and direction of the deviations from the weighted pooled regression line. The regression slope and intercepts for this pooled regression line were primarily determined by the noncoached population since the ratio of noncoached to coached individuals is approximately 4:1 in this sample. The FTC report of larger mean deviations for the coached group is therefore not unexpected. The directions or sign of the deviations would depend on the way in which the regression hyperplanes differ. The FTC analysts do not appear to have tested the critical assumptions of homogeneity of the within population regressions before they carried out and interpreted the difference in mean residuals. It would appear that any conclusions about coaching as working best for "underachievers" is not warranted based on this type of analysis.

A more satisfactory, yet still inconclusive, analysis would be centered on the coached population only, and would include the regression of the first SAT on the PSAT score and significant demographics as control variables. This analysis would be repeated for the second SAT. That is, the second SAT would be regressed on the first SAT and control demographics. The comparison of the relative sizes of the regression weights associated with the PSAT and the first SAT would indicate if a particular subgroup of the coached population benefits more from coaching.[*] The problem here is that this comparison would only be valid if one first shows that the growth rates of subsets of the coached population are the same in the absence of intervention.

---

[*] e.g., If the raw score regression weight is greater than 1.0 for the PSAT but less than 1.0 for the first SAT, then one could infer that when background variables are controlled, low scoring individuals gain more from coaching. Since we don't have reliability estimates for the PSAT and SAT for this population, such conclusions would have to be tentative.

Appendix 2
Stroud Report

Reanalysis of the Federal Trade Commission Study of Commercial Coaching for the SAT

T.W.F. STROUD

Abstract

SAT scores of students attending three commercial coaching schools are compared with those of uncoached students by means of multiple regression techniques. The techniques differ from those used in the Federal Trade Commission (FTC) reanalysis in a number of ways which are described in this report. One essential difference is that the SAT scores were predicted using a multiple regression equation based on *uncoached* students rather than pooled across coached and uncoached students. Residuals from predicted values based on this equation were then computed for *coached* students; these residuals were averaged within peak test administration months within coaching schools to produce *coaching/self-selection* effects. Bayesian methods, utilizing the concept of "borrowing strength" or "smoothing" to obtain estimates for smaller schools, were employed to estimate the overall effect at each of the three schools, averaged out over years.

The general impression given by this analysis is similar in spirit to that given by the FTC reanalysis: for the school with the strongest coaching/self-selection effects, smoothed estimates of the effects averaged over years range from 16.9 points on SAT-Math for Juniors to 28.5 points on SAT-Verbal for juniors, with standard errors in the range 9.2 to 18.6; for other schools the smoothed estimates of the effects tend to be smaller.

In exploring interactions between the coaching/self-sclection effect and various background variables, it was discovered that at one school showing overall negligible effects black students exhibited a much higher average coaching/self-selection effect (46 points) than whites. This interaction appeared to be separate from an interaction with parental income, which was also found to be statistically significant.

Summary of Background

This report is a reanalysis of data which were first analyzed by the Boston Regional Office of the Federal Trade Commission (FTC) (1978) and then reanalyzed by the Bureau of Consumer Protection (BCP) (1979) in the FTC central office in Washington, D.C. The purpose of the original study was to determine how much effect preparation under the guidance of commercial coaching schools had on the Scholastic Aptitude Test (SAT) scores and the Law School Admissions Test (LSAT) scores of the clients of those schools. The reanalysis of the SAT data was undertaken by the BCP because serious defects had been discovered in the original Boston office analysis; specifically, observational studies had been made on two groups, coached and uncoached students, and regression lines of SAT score on Preliminary Scholastic Aptitude Test (PSAT) score had been compared with no attempt either to present standard errors or to correct for values of background variables. There are, in fact, systematic differences between coached and uncoached students with respect to such background characteristics as high school achievement, race, and self-reported parental income. There are also systematic differences in the distribution of these variables for students attending the two schools included in the FTC (1979) reanalysis.

The purpose of the present study is an independent investigation of the SAT data analyzed and reanalyzed by the FTC. Extensive use is made here of background variables, as was done in the FTC reanalysis. However, this reanalysis differs from the FTC reanalysis in a number of ways. These ways will be described, following a summary of the FTC reanalysis.

The FTC data set began with about 2000 enrollees at two coaching schools in the metropolitan New York area during the period 1974-77 for whom SAT scores and background data were available. There were 600 enrollees whose names were not found in the ETS file of test-takers in certain zip code areas of Connecticut, New Jersey, and New York which were considered to be primary market areas for the participating coaching schools. Most of these, presumably, had not previously and did not subsequently take the SAT or did not take it in the metropolitan New York area. There was also a third coaching school whose enrollees' data were not analyzed by the FTC, because the number of individuals involved was considered too small. A control group of about 2500 uncoached students was chosen from among persons taking the

SAT during the same three-year period in the same geographical area using a systematic sample of every 150th student in the file provided by Educational Testing Service (ETS).

In the FTC reanalysis, multiple regression analyses were performed on six subsamples:

1. High school juniors taking the SAT for the first time in April 1975. (76 coached and 607 uncoached students)
2. High school juniors taking the SAT for the first time in April 1976. (247 coached and 617 uncoached students)
3. High school seniors taking the SAT for the second time in November 1975. (98 coached and 396 uncoached students)
4. High school seniors taking the SAT for the second time in November 1976. (177 coached and 387 uncoached students)
5. All high school students taking the SAT for the first time on any of the test dates over the 3-year period. (417 coached and 1763 uncoached students)
6. All high school students taking the SAT for the second time on any of the test dates over the 3-year period. (316 coached and 1267 uncoached students)

Background variables included in all regression analyses, in roughly decreasing order of importance, were: pretest score (PSAT Verbal or Math when SAT1, the first SAT taken, was being predicted, SAT1-Verbal or -Math when SAT2, the second SAT taken, was being predicted, using the Verbal pretest to predict the Verbal score and the Math pretest to predict the Math score), self-reported grade in English or math, self-reported rank-in-class, self-reported years of high school instruction in English or math, self-reported parental income, sex, race, high school type, and number of PSATs taken (for predicting SAT1). Coaching school (School A or B) was entered as a dummy variable. Time between pretest and test was also entered, but it did not increase prediction significantly. Only those students with complete data were entered in the regression analyses.

The analyses showed that students from School A scored significantly higher, on the average, than uncoached students. Ninety-five percent confidence limits for the differences in adjusted means, based on the median lower confidence limit and median upper confidence limit over the 12 analyses (6 subsamples, Verbal and Math scores), were 14 and 38. Coached students from School B did not earn significantly higher scores than uncoached students, with median confidence limits of -12 and +19.

The FTC (1979) report then presented a regression analysis predicting PSAT scores from the same background variables as above. When actual and predicted PSAT scores were compared, the students coached between the PSAT and SAT1 had significantly lower residual scores on the PSAT than the uncoached students. They thus appeared to be "underachievers on standardized tests." A similar analysis was carried out for SAT1 scores. The finding was that students coached between SAT1 and SAT2 had significantly lower residual scores on SAT1 than uncoached students. The conclusion, as reported in the FTC reanalysis, was that obtaining pretest scores that were lower than predicted from background variables might have been a factor in enrollees selecting themselves for coaching schools. Since pretest was the strongest predictor variable in the coaching regression analysis, this self-selection effect could account for some of the difference between adjusted means of the coached and uncoached groups.

The FTC reanalysis then attempted to control for this self-selection by recomputing the original regression analysis with the pretest removed from the set of covariates. In this analysis the adjusted means for coached students at School A were above those of uncoached students, but the difference was not significant at the 5% level (median confidence limits of -5 and +30). Finally, in an effort to see whether this self-selection in the form of underachievement on standardized tests was consistent or random, both PSAT and SAT1 scores were predicted for students who were coached between SAT1 and SAT2. Since residual scores for both PSAT and SAT1 for these coached students were found to be negative and significantly lower than residual scores for uncoached students, this was taken as evidence that the underachievement was "consistent" rather than "random." Hence, when the scores of these students improved after attendance at coaching school, the improvement was attributed to coaching rather than to any self-selection effect.

It seems quite plausible, however, especially since the coached groups are so small, that students may have selected themselves for coaching on the basis of negative measurement errors in both PSAT and SAT1. In any case, we are dealing with an observational study where the self-selection effects are likely quite complex and it is not possible, from the data provided, to disentangle the effects of self-selection from the effects of coaching. How much is a coaching effect and how much is a self-selection effect may be speculated upon, but it cannot be determined from the data.

Differences Between This Analysis and the FTC Reanalysis

In the FTC (1979) reanalysis, verbal background variables (e.g., PSAT-V, grades in English, years of English) were used to predict SAT-Verbal scores and quantitative background variables were used to predict SAT-Math scores. It was found, however, that both Verbal and Math PSAT scores were important in predicting both SAT1-V and SAT1-M. Similarly, both SAT1-V and -M scores were valuable in predicting both SAT2-V and SAT2-M for the seniors. The number of years of English taken did not additionally contribute to the predictive value of any regression equation when the other variables were already entered.

The FTC reanalysis dealt with the six subsamples given earlier. The analysis presented here deals with the peak 1975 month and peak 1976 month only—subsamples 1, 2, 3, and 4 in the FTC study—and omits the combined subsamples 5 and 6 which were regarded as too heterogeneous.

An additional difference is that missing-value techniques were used here so that students who did not report parental income, race, or rank-in-class could nevertheless be included in the present analysis; such students were excluded from the regression equations in the FTC reanalysis. Furthermore, in the FTC reanalysis, students from the coaching schools who did not receive their coaching prior to the SAT administration in question were added to the sample of uncoached students. Since this destroys the representativeness of the systematic sample of every 150th uncoached student, they were excluded in the present analysis.

Although the FTC's data set involved three coaching schools, only Schools A and B were used in the FTC reanalysis, because the enrollment in School C was considered to be too small. Furthermore, the student populations from Schools A and B were frequently amalgamated so that the report refers to "coached students." The position taken here is that each coaching school is a separate population and thus the three coaching schools are identified separately. Estimates are provided for these three schools which utilize the empirical Bayes concept of borrowing strength, and which also allow for the possibility of predicting coaching/self-selection effects in the same schools in future years (under certain assumptions).

Aims of This Analysis and Summary of Statistical Methods

Students who select a coaching program at a commercial coaching school tend to differ in two major ways from the vast majority who do not do so, namely, motivation and financial means. Those who do not have higher-than-average financial means must be even more strongly motivated. Although there are many students who do not obtain coaching who are highly motivated to do well on the SAT, there are also many whose motivation is poor. A reduced motivation was concluded to be one of the factors involved in the SAT score decline of recent years (Wirtz et al., 1977).

The view taken in the present analysis is that the data provide no way of distinguishing the effect of coaching from the effect of self-selection or motivation. There is a rough proxy for financial means (self-reported parental income) but there is no proxy for motivation. There may also be other factors in the self-selection into coaching which we have not pinpointed. Therefore, differences in SAT score after coaching between coached students and uncoached students will be called combined *coaching/self-selection effects*.

Because of the obvious differences in character between the coached and uncoached populations (e.g., higher grades and rank-in-class, higher parental income, and higher percent enrollment in private schools for coached students), it would be unreasonable to suppose that the mean SAT scores, after controlling for all covariates imaginable, would be exactly the same. Hence, an hypothesis of no coaching/self-selection effect was avoided. Instead, the coaching/self-selection effect at each coaching school was estimated along with a standard error for the estimate. A coaching/self-selection effect whose magnitude exceeds a certain number of standard errors (say 2 or 2.5) would be regarded as being "significant" in the traditional sense.

The analyses contained herein were performed separately on four dependent variables: SAT1-V scores of juniors, SAT1-M scores of juniors, SAT2-V scores of seniors, and SAT2-M scores of seniors. As indicated earlier, analysis was restricted to students whose SAT was taken in the peak month (April for juniors, November for seniors) of 1975 and 1976.

The first stage of the analysis was to obtain a multiple regression equation for juniors and a multiple regression equation for seniors for each dependent variable based on the sample of uncoached students. Next, these equations were applied to the coached students in order to predict the SAT scores for the coached students as if they had been uncoached students with the same values of predictor variables. Residuals from these regressions were calculated. The coaching/self-selection effect is thought of as the typical value of these residuals. Separate coaching/self-selection effects, together with standard errors, were calculated for each of the two administration years for each of the three coaching schools. The third and final stage of the analysis was an empirical Bayes smoothing of the effects across administration years and coaching schools. The analysis is described in greater detail in the following three sections.

The basic aim of these analyses was to estimate the magnitude of the coaching/self-selection effects. If these combined effects are small, this puts a ceiling on the effects due to coaching at the three schools considered. If these combined effects are large, one can only speculate on how much of the effects can be attributed to coaching and how much to self-selection.

Regression Equations for the Uncoached Students

The model for the data for the uncoached students is

$$Y_i = \beta_0 + \Sigma_{j=1}^{p} \beta_j X_{ij} + e_i \quad , \qquad i = 1,2,\ldots,N \qquad (1)$$

where Y_i represents the score on the dependent variable SAT1-V, SAT1-M, SAT2-V, or SAT2-M of the *ith* student (junior or senior, depending on the dependent variable), N is the number of uncoached students, and X_{ij} represents the score on the *jth* predictor variable of the *ith* student. The set of p predictor variables includes a dummy variable representing the administration year. The error e_i is assumed to have constant variance σ^2.

Although the Y_i were measured for all students, some students had X_{ij} missing for one or more j. Thus, instead of obtaining the coefficients β_j in equation (1) from a least-squares regression program, the following method was used. The maximum-likelihood estimates of the means, variances, and covariances of $(Y_i, X_{i1}, X_{i2}, \ldots, X_{ip})$ were computed using the program based on a multivariate normal model for all variables with some observations missing (Rubin, 1974). Estimates of the coefficients β_0 and β_j $(j = 1,\ldots,p)$ were obtained from the resulting estimates using the sweep operator (SWP) on all predictor variables. If all variables were observed on all individuals, the estimated βs so obtained would be the least-squares estimates. This method is more efficient than that of computing covariances using pairwise-observed data and provides consistent estimates under more modest conditions (Rubin, 1976).

Regression coefficients were also computed for use with coached students who had one or more of the predictor variables of race, high school rank, and parental income missing. For example, for a student who had all three of these variables missing a regression equation which did not involve these three predictors was used. To obtain this equation, one begins with the matrix of maximum-likelihood estimates referred to in the preceding paragraph and applies the SWP operator to the set of all predictor variables excluding race, rank, and parental income. This yields the regression coefficients presented in the last column of Tables 1a to 1d, which show all regression coefficients for predicting SAT1-V, SAT1-M, SAT2-V, and SAT2-M, respectively. Columns 2 to 7 give regression coefficients for other missing variable patterns, and column 1 gives the complete-data regression. The list of all predictor variables studied appears in Table 1e.

Estimates and Standard Errors of the Coaching/Self-Selection Effect

The next step is to predict the SAT scores of coached students using the equation (1) which was developed for uncoached students. Denote by $b_0, b_1, \ldots, b_p$ the estimates of $\beta_0, \beta_1, \ldots, \beta_p$ obtained from that analysis. Define

$$\hat{Y}_{sti} = b_0 + \Sigma_{j=1}^{p} b_j X_{stij} \qquad (2)$$

for the *ith* student from coaching school s taking the SAT in year t whose background data $\{X_{sti1}, X_{sti2}, \ldots, X_{stip}\}$ is complete. For a student with missing data, background variables comprise a smaller set of predictor variables having subscripts in the subset $A \subset \{1,2,\ldots,p\}$ (and whose scores are missing in the complementary set A^c), define

$$\hat{Y}_{sti} = b_0^A + \sum_{j \epsilon A} b_j^A X_{stij} \qquad (3)$$

where b_j^A are the regression coefficient estimates based on an equation using only the variables in the set A. The values of b_j and b_j^A for the subsets A being considered are the values presented in Table 1a to 1e. Y_{sti} represents the predicted score of the (s,t,i)*th* coached student, ignoring the fact that he was coached.

For each coached student, the quantity $d_{sti} = Y_{sti} - \hat{Y}_{sti}$ is then computed. This is the residual of his actual score from his predicted score and represents the estimated coaching/self-selection effect for that one student. The average of these residuals within the (s,t) cell, denoted by $\bar{d}_{st}$, is an unbiased estimate of the parameter $\bar{\delta}_{st}$, which is defined as the average difference between the value of E(Y|X) under a "correct" model for coached students in the (s,t) cell (which may be a different value for each student in the cell) and the value of E(Y|X) under the uncoached model (1). If one assumed a strictly additive coaching/self-selection effect according to the model

$$Y_{sti} = \beta_0 + \Sigma_{j=1}^{p} \beta_j X_{stij} + \delta_{st} + e_{sti} \quad , \tag{4}$$

where β_0, $\beta_1,\ldots,\beta_p$ are the same parameters as in the uncoached model, then it would be better to use, instead of $\bar{d}_{st}$, a weighted average which would be minimum-variance unbiased under this model. However, as pointed out in Cochran and Rubin (1973, p. 424), $\bar{d}_{st}$ is an unbiased estimate of $\bar{\delta}_{st}$ even when $\beta_0,\beta_1,\ldots,\beta_p$ take on different values in the coached model.

Tables 2 and 3 show the sample sizes for the uncoached students and the coached students, respectively. Table 3 also shows the number of students who were removed from the analysis of coached students because of rare or extensive missing data patterns. To have included these data patterns in the analysis would have made the computing task much more onerous and was not deemed to be worthwhile.

Table 4 shows the estimated coaching/self-selection effects $\bar{d}_{st}$, together with estimated standard errors. These standard errors are calculated on the assumption that model (3) holds for the coached students.

Smoothed Estimates for the Three Coaching Schools and Predictive Distributions for a Future Year

Bayesian methods are recommended to predict the coaching/self-selection effect in a future year. The strongest advantage of Bayesian prediction is that a *predictive distribution* for a future value can be obtained, which can be made to incorporate uncertainty due to the effects of sampling in the observed data, uncertainty due to year-to-year fluctuation, and uncertainty due to not knowing certain quantities such as variances between schools and variances between years. An additional advantage is that information from schools with large numbers of students can be used to improve the estimates for schools with smaller numbers of students. (In problems involving larger numbers of groups, a general improvement can be brought about by smoothing all predictions toward the grand mean.) In this section the Bayesian analysis is incorporated into the model and the smoothed estimates of the Verbal and Math coaching/self-selection effects for each school averaged over years is presented. These smoothed estimates also serve as point predictors for a future year. Standard errors for these predictions are also shown. The validity of these predictions and standard errors depend upon the following assumptions:

1. The observed d_{st} are unbiased estimates of the coaching/self-selection effects δ_{st}; the standard errors shown in Table 4 represent the uncertainty of these estimates.
2. The quantities δ_{st} within school s for different t, both past and future, are normally distributed about the expected effect δ_s^* for that school, which remains constant within the future range being predicted. The variance of the δ_{st} about δ_s^* is the same for each school s. The value being predicted is δ_{st} where t represents a future year.

In the Bayesian analysis, the prior distribution of the expected effects δ_s^* is a normal distribution. The standard deviation of this distribution and the standard deviation of the δ_{st} within school have independent prior distributions which are each approximately constant over the range of plausible values. The posterior means and standard deviations of the δ_s^* are shown in Table 5. The means and standard deviations of the predictive distributions for each δ_{st} (s = A, B, C; t = a future year) for the four dependent variables (juniors' SAT1-V and SAT1-M, seniors' SAT2-V and SAT2-M) are shown in Table 6.

Interactions between Coaching/Self-Selection Effects and Background Variables

The question now arises: is the coaching/self-selection effect greater for some subsets of the student population than for others? To explore possible evidence of such interactions, a multiple regression analysis of the Verbal and of the Math individual coaching/self-selection effects was performed on the set of variables which had been used to predict the SAT scores. This was done for the three largest cells of coached students, namely the

SAT2 seniors of School A and the SAT1 juniors of Schools A and B, all from the 1976 peak month administration. Only students with no missing values on background variables were entered into this exploratory analysis; the sample sizes were 102, 103, and 85, respectively. For the two School A cells, the sums of squares and degrees of freedom for multiple regression and for residual are shown in Table 7; the F-values of 1.106 (Verbal) and 1.119 (Math) for seniors and the F-values of 0.606 (Verbal) and 0.735 (Math) for juniors are not significant.

For the School B juniors, the sums of squares and degrees of freedom for multiple regression and for residual are also shown in Table 7; the F-value of 2.112 (Verbal) is significant at the .05 level, while the F-value of 1.587 (Math) is not. To discover which variables contributed most to the analysis of the Verbal individual coaching/self-selection effects, the squared correlation coefficient (r^2) was calculated between this dependent variable and each of the predictor variables; these are shown in Table 8. The most significant predictor variable is RACE2 (= 1 for black, 0 for nonblack) with $r^2 = .0830$. The square of the partial correlation coefficient (r^2_{RACE2}) was calculated between every other predictor variable and the Verbal individual coaching/self-selection effect and is also shown in Table 8. The most important predictor variable after RACE2 is entered is INCOME, with $r^2_{RACE2} = .0664$, or $r_{RACE2} = -.258$. This differs little from the $r^2 = .0626$ ($r = -.250$) ignoring RACE2, indicating that the effect of self-reported parental income and the effect of black vs. nonblack are essentially separate from each other. Both these effects are statistically significant at the .05 level.

The remainder of this section is devoted to the study of differences in the coaching/self-selection effect across races or minority groups.

Table 9 shows values of background variables and of the individual coaching/self-selection effects for coached students who identified themselves as blacks who took SAT1 as a junior or who took SAT2 as a senior during the peak administration month (and who were thus included in the analysis of this report). There is a strong tendency for both the Verbal and Math coaching/self-selection effects to be above the average (cf. Table 4). For the Verbal effects the pattern is striking: thirteen out of fifteen effects are above average and the remaining two are only very slightly below average.

A significance test was performed between the blacks and whites from Coaching School B, since 13 out of 15 coached blacks were coached at that school. The variable studied was the Verbal individual coaching/self-selection effect; the null hypothesis was that the mean for blacks equals the mean for whites. Since the overall Verbal averages for the three School B cells (1976 juniors, 1975 seniors, 1976 seniors) were within a fairly narrow range (4.16 to 11.12, from Table 4), the students from these cells were pooled.

The results of this black-white comparison are shown in Table 10. (The data set used excludes 18 students who identified themselves as belonging to other nonblack minority groups and who are among those listed in Table 12.) Equality of population variances for black and white coaching/self-selection effects was not assumed. The Welch approximate t-test was used (see, e.g., Brownlee, 1965, p. 299). The 2-sided test showed significance at the .001 level, with the average for blacks being 46.7 points above that for whites. The average Math coaching/self-selection effect for blacks also exceeded that for whites, but the difference was not statistically significant.

There were five black juniors taking SAT1 and one black senior taking SAT2 at times other than peak administration months, so their scores have not been included in any analysis described so far. A tabulation of their data appears in Table 11. The Verbal effects are higher than the corresponding cell averages in five cases out of six; the average of these six Verbal effects is 52.6 points above the average of the corresponding cell average values. The Math effects are higher than the cell averages four times out of six, averaging 13.6 points above cell average. This pattern is virtually the same as that observed for the blacks who took the SAT in the peak administration months.

Finally, the data for students who identified themselves as belonging to minority groups other than black are presented in Table 12. Both peak-month and nonpeak-month SAT-takers are included. The individual coaching/self-selection effects bear more resemblance to those of white students than to those of black students.

Conclusions

The year-by-year estimated effects for coaching schools A and B, as given in Table 4, are of a similar size to those presented in the FTC report (1979), in that for School A the coaching/self-selection effects range from 20 to 34 SAT points (from 3 to 6.9 standard errors) and for School B the effects are nonsignificant except for the seniors' SAT2-M in 1975 (26 points = 2.38 standard errors). For School C the estimates fluctuate markedly from year to year, the strongest effect being the juniors' SAT1-M in 1975, +2.2 standard errors as compared with -1.5 standard errors in 1976.

The smoothed estimates of the school-by-school effects averaged over years, given by Table 5, are only slightly smaller for School A than the unsmoothed estimates. However, the standard deviations shown in Table 5 are larger than the within-year standard errors of Table 4, reflecting the year-to-year fluctuation exhibited in the data. For a future year one would predict a positive coaching/self-selection effect at all schools (Table 6), but with a standard error greater than the value of the predicted effect except for SAT-V at School A (predicted effect = 1.8 standard errors for juniors, 1.3 for seniors) and SAT-V for juniors at School C (predicted effect = 1.3 standard errors).

It is impossible to judge the extent to which each of these effects is due to the coaching program or rather to personal characteristics underlying self-selection of the coached students. However, in a randomized experiment (Alderman and Powers, 1979; Rubin, 1979) which was designed to be free of self-selection effects, special preparation programs offered in eight high schools had estimated coaching effects ranging from 6.3 to 12.5 SAT points, or 0.9 to 1.9 standard errors. The larger estimates derived from the data analyzed in this report are consistent with the possibility that both the motivation to do well on tests, as exemplified by enrollment in a commercial coaching school, and the coaching program itself contribute to a small gain in SAT scores.

There is a clear and strong positive difference in the coaching/self-selection effect of blacks over whites at School B, as shown by Table 10. Whereas the average effect for the 132 whites analyzed here was virtually nil, the average coaching/self-selection effect for the 13 blacks was 46 points, or 4.47 standard errors ($S_x/\sqrt{N} = 10.39$). One possible interpretation of this phenomenon is that blacks have an initial disadvantage in test-taking that can be overcome by coaching. However, it is also possible that there is a greater difference in personal factors such as motivation between blacks who enroll in a coaching school and blacks who do not than between whites who enroll and whites who do not. We caution that it would be unwarranted to take the relative unimportance of the coefficients for race in the multiple regressions (Table 1) as evidence against the hypothesis that blacks are disadvantaged in test-taking. The strongest predictor variable in the multiple regression is the pretest (PSAT or SAT1), and a race effect could conceivably be operating there.

To summarize, this study has shown certain relatively small but unmistakable effects of coaching and/or of self-selection. Further study is needed, perhaps in the form of additional randomized experiments, to shed more light on the subject of what portion of these effects is due to coaching.

Table 1a

Regression Coefficients for Predicting SAT I-V Scores for Juniors
(April Administration)
For Complete Data and Various Combinations of Missing Variables
(For names of predictor variables see Table 1e)

	Equation Number							
Variable No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
0	-113	-136	-112	-136	-109	-133	-108	-132
1	.37	-.19	.33	-.23	.99	.44	.97	.41
2	1.692	1.908	1.680	1.896	1.745	1.958	1.730	1.945
3	-.033	.181	-.042	.173	-.052	.161	-.069	.146
4	-3.14	-3.99	-3.04	-3.87	-3.77	-4.61	-3.59	-4.42
5	.4817	.4786	.4958	.4940	—	—	—	—
6	-6.1	-6.5	—	—	-8.5	-8.9	—	—
7	-1.1	-1.4	—	—	-3.8	-4.0	—	—
8	.3131	—	.3151	—	.3107	—	.3136	—
9	13.3	14.1	13.4	14.2	13.4	14.2	13.6	14.4
10	.174	.199	.182	.208	.232	.257	.247	.273
11	1.093	1.197	1.096	1.201	1.142	1.245	1.153	1.258
12	7.122	7.217	7.126	7.222	7.140	7.235	7.148	7.244

Table 1b

Regression Coefficients for Predicting SAT I-M Scores for Juniors
(April Administration)
For Complete Data and Various Combinations of Missing Variables
(For names of predictor variables see Table 1e)

	Equation Number							
Variable No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
0	-61.6	-85.7	-65.0	-88.9	-60.2	-84.2	-63.5	-87.4
1	-8.43	-8.99	-8.43	-8.98	-8.19	-8.75	-8.19	-8.74
2	.028	.246	.063	.278	.049	.265	.082	.296
3	1.575	1.792	1.571	1.785	1.567	1.784	1.561	1.775
4	-8.67	-9.52	-8.39	-9.23	-8.92	-9.76	-8.60	-9.43
5	.1889	.1858	.1870	.1853	—	—	—	—
6	18.1	17.6	—	—	17.1	16.1	—	—
7	-16.8	-17.1	—	—	-17.8	-18.1	—	—
8	.3163	—	.3135	—	.3153	—	.3130	—
9	-14.1	-13.4	-14.4	-13.7	-14.1	-13.3	-14.4	-13.6
10	7.984	8.009	7.968	7.993	8.006	8.031	7.992	8.018
11	6.871	6.976	6.904	7.009	6.890	6.994	6.926	7.030
12	1.478	1.575	1.474	1.571	1.485	1.582	1.483	1.579

Table 1c

Regression Coefficients for Predicting SAT2-V Scores for Seniors
(November Administration)
For Complete Data and Various Combinations of Missing Variables

(For names of predictor variables see Table 1e)

	Equation Number							
Variable No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
0	-24.3	-24.2	-26.8	-26.7	-27.6	-28.3	-30.1	-30.7
1	8.53	8.53	8.51	8.51	8.20	8.19	8.17	8.16
2	.568	.568	.590	.589	.586	.591	.609	.613
3	-.128	-.129	-.149	-.151	-.148	-.141	-.167	-.160
4	-.56	-.55	-.14	-.13	-.17	-.19	.23	.21
5	-.260	-.260	-.256	-.256	—	—	—	—
6	2.3	2.3	—	—	4.0	4.0	—	—
7	-13.2	-13.2	—	—	-12.5	-12.5	—	—
8	-.001	—	-.002	—	.008	—	.008	—
9	4.99	4.98	4.85	4.85	5.03	5.05	4.86	4.88
10	-.974	-.973	-1.04	-1.04	-.767	-.775	-.843	-.851
11	.709	.708	.730	.729	.675	.678	.695	.697
12	3.358	3.358	3.363	3.362	3.375	3.377	3.379	3.381
13	.5706	.5706	.5698	.5698	.5651	.5652	.5643	.5643
14	.0110	.0110	.0141	.0141	.0111	.0111	.0141	.0141

Table 1d

Regression Coefficients for Predicting SAT2-M Scores for Seniors
(November Administration)
For Complete Data and Various Combinations of Missing Variables

(For names of predictor variables see Table 1e)

	Equation Number							
Variable No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
0	-8.8	-14.3	-12.2	-17.7	-11.9	-18.2	-15.2	-21.4
1	-2.59	-2.65	-2.61	-2.67	-2.90	-2.97	-2.92	-3.00
2	-.356	-.313	-.327	-.284	-.340	-.291	-.310	-.262
3	.686	.746	.656	.715	.668	.734	.640	.706
4	-9.33	-9.53	-8.73	-8.94	-8.96	-9.19	-8.40	-8.63
5	-.240	-.246	-.232	-.237	—	—	—	—
6	2.2	2.2	—	—	3.8	3.8	—	—
7	-18.3	-18.3	—	—	-17.7	-17.7	—	—
8	.074	—	.073	—	.083	—	.082	—
9	-9.37	-9.16	-9.53	-9.33	-9.32	-9.10	-9.52	-9.30
10	8.627	8.55	8.54	8.46	8.82	8.74	8.72	8.64
11	3.652	3.675	3.681	3.704	3.621	3.646	3.650	3.674
12	-.026	-.010	-.019	-.003	-.010	-.008	-.004	.014
13	.0972	.0980	.0960	.0969	.0920	.0929	.0910	.0918
14	.5329	.5332	.5371	.5374	.5329	.5333	.5371	.5375

Table 1e

Predictor Variables Used in Regression Equations

Number	*Variable Name*
0	y-Intercept
1	Administration year (0 = 1975, 1 = 1976)
2	Self-reported English grade (out of 100)
3	Self-reported Math grade (out of 100)
4	High School type (0 = public, 1 = other)
5	Self-reported parental income (in thousands)
6	RACE1 (1 if Oriental, Native, etc.; 0 if otherwise)
7	RACE2 (1 if black, 0 if otherwise)
8	Self-reported rank (scaled out of 100)
9	Sex
10	Years of mathematics
11	PSAT-M
12	PSAT-V
13	SAT1-V
14	SAT1-M

Table 2

Numbers of Uncoached Students With Complete Data and With Various Missing Data Patterns

(1975 and 1976 combined, peak administration months)

Data Pattern	*Number of Juniors Taking SAT1*	*Number of Seniors Taking SAT2*
Complete	711	517
Missing: Income, Race, Rank	1	1
Income, Race	9	6
Income, Rank	12	6
Income	143	94
Race, Rank	0	1
Race	10	7
Rank	19	15
Other Background Variables	56	37
PSAT	0	67
PSAT & Background	0	17
Total Students Used in Analysis	961	768

(In analysis of juniors, only students with PSAT scores were included.)

Table 3

Numbers of Coached Juniors With Complete and With Incomplete Data

Data Pattern		*School A*		*School B*	*School C*	
		1975	*1976*	*1976*	*1975*	*1976*
Complete		49	103	85	16	9
Missing:	Income, Race, Rank	0	2	1	0	0
	Income, Race	3	3	0	1	0
	Income, Rank	3	1	2	0	0
	Income	14	19	12	3	2
	Race, Rank	0	1	0	0	0
	Race	1	4	2	1	1
	Rank	1	1	6	0	1
Total Students Used in Analysis		71	134	108	21	13
Other Data Patterns (not used in analysis)		5	4	3	3	1
Total Coached Students with PSAT and some background data		76	138	111	24	14

Numbers of Coached Seniors With Complete and With Incomplete Data

Data Pattern		*School A*		*School B*		*School C*	
		1975	*1976*	*1975*	*1976*	*1975*	*1976*
Complete		45	102	15	24	7	4
Missing:	Income, Race, Rank	0	0	0	0	0	0
	Income, Race	2	1	0	0	0	0
	Income, Rank	2	2	1	0	0	0
	Income	10	13	2	9	1	2
	Race, Rank	1	0	1	0	0	0
	Race	5	3	0	0	0	0
	Rank	0	2	2	1	0	1
Total Students Used in Analysis		65	123	21	34	9	5
Other Data Patterns (not used in analysis)		5	7	0	0	1	0
Total Coached Students with PSAT and some background data		70	130	21	34	10	5

Table 4
Coaching/Self-Selection Effects and Their Standard Errors

	SAT-VERBAL					SAT-MATH				
	1975 SAT1-V			1976 SAT1-V		1975 SAT1-M			1976 SAT1-M	
	School	*Effect*	*S E*	*Effect*	*S E*	*School*	*Effect*	*S E*	*Effect*	*S E*
	A	27.44	6.66	34.13	5.23	A	22.36	7.45	20.37	5.85
Juniors	B	—	—	4.16	6.38	B	—	—	3.71	7.14
	C	17.88	11.19	25.64	14.10	C	27.68	12.52	-23.32	15.79
	1975 SAT2-V			1976 SAT2-V		1975 SAT2-M			1976 SAT2-M	
	School	*Effect*	*S E*	*Effect*	*S E*	*School*	*Effect*	*S E*	*Effect*	*S E*
	A	32.07	5.86	31.34	4.51	A	22.54	6.41	32.62	4.93
Seniors	B	11.12	10.16	4.76	8.05	B	26.40	11.10	8.07	8.80
	C	-10.47	14.02	20.30	18.70	C	24.34	15.33	-21.57	20.44

Table 5
Smoothed Estimates of Coaching/Self-Selection Effects
(Posterior Means and Standard Deviations)

JUNIORS TAKING SAT1

	Verbal		Quantitative	
School	*Mean*	*Std Deviation*	*Mean*	*Std Deviation*
A	28.5	9.2	16.9	18.6
B	9.7	12.8	7.8	23.0
C	20.9	10.7	8.4	19.5

SENIORS TAKING SAT2

	Verbal		Quantitative	
School	*Mean*	*Std Deviation*	*Mean*	*Std Deviation*
A	28.4	12.1	24.0	15.1
B	9.4	12.3	17.2	15.3
C	6.1	15.0	12.0	17.6

Table 6

Predictive Means and Standard Standard Deviations of Coaching/Self-Selection Effects for a Future Year

JUNIORS TAKING SAT1

	Verbal		Quantitative	
School	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*
A	28.5	15.7	16.9	35.9
B	9.7	18.0	7.8	38.4
C	20.9	16.7	8.4	36.4

SENIORS TAKING SAT2

	Verbal		Quantitative	
School	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*
A	28.4	21.6	24.0	28.7
B	9.4	21.8	17.2	28.8
C	6.1	23.4	12.0	30.1

Table 7

Multiple Regression Statistics for the Regression of the Individual Coaching/Self-Selection Effect on Pretest and Background Variables (1976 Peak Month Administration)

	School A Juniors		*School A Seniors*		*School B Juniors*	
	SAT1-V	SAT1-Q	SAT2-V	SAT2-Q	SAT1-V	SAT1-Q
Regression: SS	21396	23353	27461	32367	61005	40070
Regression: DF	10*	10*	13	13	11	11
Residual: SS	324662	292338	168088	195754	191707	167531
Residual: DF	92	92	88	88	73	73
R^2	.062	.074	.140	.142	.241	.193
R	.249	.272	.374	.377	.481	.439
F	.606	.735	1.106	1.119	2.112	1.587

---

* The predictor variable RACE2 (black vs nonblack) was omitted since there were no black students in this cell

Table 8

Squared Correlation Coefficients between the Individual Coaching/Self-Selection Effects and Each Predictor Variable: Zero-Order Correlation (r^2) and Partial Correlation after Entering RACE2 (r^2_{RACE2})

(School B Juniors, April 1976 (Peak) Administration; Verbal Only)

Predictor Variable	r^2	r^2_{RACE2}
Self-reported English grade	.0001	.0005
Self-reported Math grade	.0006	.0002
High school type	.0598	.0524
Self-reported parental income	.0626	.0664
RACE1	.0304	.0429
RACE2	.0830	—
Self-reported rank	.0168	.0093
Sex	.0492	.0433
Years of mathematics	.0121	.0140
PSAT-V	.0211	.0282
PSAT-M	.0001	.0004

Table 9

Background Variables and Individual Coaching/Self-Selection Effects for Coached Students Identified as Black

	Coaching School	Adm Date	English Grade	Math Grade	H.S. Type*	Potential Income	Rank in Class	Sex	Years of Math	PSAT		SAT1		SAT2		C/SS Effect	
										V	M	V	M	V	M	V	M
Juniors	B	4/76	85	95	NP	7	70	F	4	39	33	390	380			-0.7	12.0
	B	4/76	85	75	NP	—	70	F	2	40	42	460	450			48.4	64.7
	B	4/76	85	85	NP	5	70	M	4	31	43	400	360			69.3	-62.9
	B	4/76	85	95	P	17	70	M	4	49	49	610	560			136.0	42.6
	B	4/76	85	85	P	10	—	M	4	49	36	480	400			24.0	-10.2
	B	4/76	85	95	P	—	70	F	4	33	58	450	550			66.9	8.6
	B	4/76	95	75	NP	35	85	F	2	45	43	500	500			20.2	91.5
	B	4/76	95	85	NP	35	85	F	3	53	40	610	510			76.7	86.6
	C	4/76	85	85	NP	10	70	M	3	31	31	390	310			70.2	-23.3
Seniors	A	11/76	85	95	P	29	95	M	4	35	46	400	460	420	560	35.9	78.5
	B	11/75	85	85	NP	8	—	M	2	38	48	460	580	500	560	69.2	29.0
	B	11/76	75	85	NP	—	30	F	4	41	38	440	390	460	360	36.5	-34.5
	B	11/76	85	85	NP	35	50	M	4	32	34	290	350	340	330	39.3	-17.5
	B	11/76	85	85	NP	5	70	F	4	35	29	290	250	320	280	1.1	4.8
	B	11/76	85	85	NP	—	50	F	3	33	33	270	280	320	390	17.4	98.4

---

*—P = Public, NP = Nonpublic

Table 10

Test of Equality of Mean Verbal Coaching/Self-Selection Effect for Blacks vs. Whites for Coaching School B

(Peak Administration Months, 1975 and 1976, Junior SAT1 and Senior SAT2 pooled)

	Blacks x	*Whites y*
Sample Size	$N_x = 13$	$N_y = 132$
Average Effect	$\bar{X} = 46.494$	$\bar{Y} = -.255$
Sample Variance	$s_X^2 = 1404.7$	$s_Y^2 = 2370.5$

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{N_x} + \frac{s_Y^2}{N_y}}} = 4.165$$

Degrees of Freedom (Welch) = 16.28

$P < .001$ (two-sided)

$t_{.0005} = 4.00$

Table 11

Data for Six Black Coached Students Who Took the SAT at Times Other Than Peak Administrations

Junior (J) or Senior (S)	*J*	*J*	*J*	*J*	*J*	*S*
Administration Date	6/75	12/75	6/76	12/76	6/75	6/75
Coaching School	A	B	B	B	C	A
English Grade	87	85	75	85	95	95
Math Grade	85	95	85	95	85	85
High School Type	0	1	1	0	0	0
Parental Income	10	19	—	29	35	8
Rank in Class	85	85	70	85	70	50
Sex	1	1	2	1	1	2
Years of Math	3	3	4	4	3	1
PSAT-V	29	32	39	33	54	33
PSAT-M	41	41	43	38	47	34
SAT1-V	410	450	400	410	550	280
SAT1-M	490	380	430	470	490	360
SAT2-V	—	—	—	—	—	450
SAT2-M	—	—	—	—	—	430
C/SS Effect (V)[1]	82.7	103.4	11.9	51.5	17.2	138.8
Average C/SS Effect (V)[2]	27.4	4.2	4.2	4.2	17.9	32.1
C/SS Effect (M)[1]	68.9	-45.8	8.1	44.8	-9.4	98.6
Average C/SS Effect (M)[2]	22.4	3.7	3.7	3.7	27.7	22.5

---

[1] Calculated from regression equation based on nearest Peak Administration Date

[2] Average V- or M- effect for Coaching School at nearest Peak Administration Date

Table 12

Background Variables and Individual Coaching/Self-Selection Effects for Coached Students Identified as Nonblack Minority Group Members

	Coaching School	Adm Date	English Grade	Math Grade	H S Type*	Potential Income	Rank in Class	Sex	Years of Math	PSAT		SAT1		SAT2		C/SS Effect	
										V	M	V	M	V	M	V	M
Asian American Juniors	A	4/76	85	95	P	35	95	M	4	40	54	580	530			153.1	-54.6
	B	4/76	95	95	P	10	85	F	5	46	73	580	690			74.3	-20.3
	B	4/76	95	95	P	—	—	M	4	30	51	450	490			93.0	-51.3
	B	4/76	95	95	P	10	95	M	5	40	66	470	680			24.9	9.4
Seniors	A	11/76	75	95	NP	35	50	M	5	44	71	500	710	500	760	1.4	25.9
	A	11/76	95	95	P	23	95	M	5	55	50	610	710	620	770	21.6	93.7
	A	11/75	85	95	P	21	85	M	3	30	45	300	520	450	550	129.8	34.0
	B	11/75	85	85	NP	25	70	M	2	50	41	530	510	540	520	23.6	29.1
	C	11/76	75	95	NP	35	85	M	5	49	51	520	610	510	560	-1.5	-52.5
Puerto Rican Juniors	A	4/76	85	95	NP	13	30	F	4	29	34	330	400			24.0	16.6
	B	4/76	75	75	NP	2	50	M	3	40	38	380	360			20.1	-45.8
	B	12/75	85	85	NP	—	70	M	3	38	35	420	450			49.1	43.2
	B	4/76	95	85	NP	5	70	F	4	44	41	540	480			93.6	30.6
	B	12/76	85	65	NP	7	50	F	3	43	35	380	450			-30.9	89.0
	B	4/76	85	75	NP	35	50	M	3	35	34	300	410			-52.7	32.6
	B	4/76	85	85	NP	10	—	M	3	41	37	430	420			39.9	-1.7
	B	6/76	95	95	NP	5	85	F	4	44	43	440	440			-12.9	-43.6
	B	6/76	95	85	NP	7	85	F	4	37	37	390	400			-7.8	-16.6
Seniors	B	11/75	95	85	P	16	50	F	4	34	46	440	450	380	430	-45.8	-53.2
	B	11/76	95	95	NP	—	95	F	4	44	39	320	310	350	400	-42.8	28.9
	B	11/76	95	95	NP	14	95	F	3	43	53	440	530	460	570	-43.1	-49.1
	B	11/76	85	85	NP	17	70	F	3	27	31	310	360	400	350	-10.0	28.7
	B	11/75	95	95	NP	16	85	M	3	41	50	390	510	350	550	-68.4	24.4
Other Nonblack Juniors	A	4/76	87	85	P	—	85	F	5	48	53	480	490			-8.7	-71.4
	A	4/76	95	95	P	25	95	M	4	41	55	460	630			12.7	38.6
	A	4/76	95	85	NP	29	70	M	4	45	53	540	560			75.2	8.0
	A	4/76	95	85	P	35	70	F	3	42	46	480	550			25.0	62.9
	A	4/76	95	95	P	16	95	M	4	43	30	460	330			30.2	-90.8
	A	4/75	95	95	P	35	95	F	3	57	54	570	650			-7.7	53.7
	A	4/76	85	85	NP	13	70	M	3	32	43	360	420			23.5	32.8
	A	4/75	95	85	P	35	70	M	2	37	33	480	450			88.7	45.0
	A	4/75	95	85	P	3	95	F	3	52	50	690	540			165.0	-0.6
	B	4/76	95	95	NP	7	—	F	3	43	55	490	530			30.4	-26.9
	B	4/76	95	95	NP	2[illegible]	85	F	3	44	51	450	510			-19.2	-23.6
	B	4/76	85	75	NP	7	50	F	3	44	45	450	440			21.3	-6.9
	B	4/76	85	75	NP	21	50	M	4	20	24	260	240			31.6	-51.9
	C	4/76	95	95	P	16	95	M	4	49	48	490	480			-2.2	-73.4
Seniors	A	11/76	85	95	P	25	85	M	5	40	70	400	630	400	700	35.3	11.0
	A	11/76	85	85	NP	14	—	M	4	37	51	450	480	450	560	6.8	38.8
	A	11/75	95	95	NP	17	70	M	3	44	58	470	600	560	660	79.5	50.8
	A	12/75	85	85	NP	5	70	F	3	38	46	420	480	430	520	4.3	33.1
	B	11/76	85	95	NP	—	70	F	4	24	30	230	380	220	320	80.5	-2.3
	B	12/76	75	85	NP	13	50	F	2	20	32	260	270	250	310	-13.5	12.3

---

*—P = Public, NP = Nonpublic

References

Alderman, D. L., & Powers, D. E. *The effects of special preparation on SAT-Verbal scores.* Research Report 79-1. Princeton, NJ: Educational Testing Service, 1979.

Brownlee, K. A. *Statistical theory and methodology*, 2nd ed. New York: J. Wiley & Sons, 1965.

Cochran. W. G., & Rubin, D. B. Controlling bias in observational studies: A review. *Sankhya A*, 19/3, **35**(4), 417-446.

Federal Trade Commission, Boston Office, 1978.

Federal Trade Commission, Washington, DC, 1979.

Rubin, D. B. Characterizing the estimation of parameters in incomplete-data problems. *Journal of the American Statistical Association*, 1974, **69**, 467-474.

Rubin, D. B. Inference and missing data. *Biometrika*, 1976, **63**(3), 581-592.

Rubin, D. B. *Estimation in parallel randomized experiments.* Research Bulletin. Princeton, NJ: Educational Testing Service, 1979.

Wirtz, W. & Associates. *On further examination.* New York: College Board, 1977.

Appendix 3
Rock Report

Disentangling Coaching Effects and Differential Growth in the FTC Commercial Coaching Study

DONALD A. ROCK

The FTC study is a pretest/posttest design where subjects are not randomly assigned to groups. In the nomenclature of Campbell and Stanley (1963), we have a special case of a quasi-experimental design known as the nonequivalent control-group design. In the classic experiment with random assignment the only rival explanation to the treatment being the causal factor is sampling error. However, for the nonequivalent control-group design, both sampling error and possible selection bias are plausible rival hypotheses.

The FTC Boston Regional Office analysis (1978) and the FTC Bureau of Consumer Protection reanalysis (1979) relied heavily on the ability of the traditional analysis of covariance model (ANCOVA) to rule out rival hypotheses. In order to accept the FTC's interpretation of their ANCOVA results as providing reasonable estimates of the treatment effect (coaching), rather strong assumptions have to be made about the appropriateness and completeness of their covariance adjustment model for the data set. The reanalysis presented in this paper will show in some detail that the verbal data do not support the strong assumptions required to arrive at their point estimates of the verbal coaching effect. The mathematics data will be shown to be more consistent with their specification of the analysis of covariance model and thus more likely to produce reasonable estimates of the math coaching effects. An adjustment procedure appropriate for the verbal data will be outlined and applied to the data. New estimates of the coaching effect will be computed and compared with the results based on the standard ANCOVA model.

The statistical design and data base used in the FTC analysis can best be depicted in the Campbell & Stanley framework as follows:

Treatment (coaching)	O_1	I	O_2
Control	O_1		O_2
	t_1		t_2

where the O's are observations on the treatment and control subjects at two different points in time, t_1 and t_2. The I indicates a formal intervention has taken place for the treatment group at a point in time between the first and second observation. The ANCOVA model traditionally used in estimating the intervention effect in this type of experiment can be formulated in terms of the general linear model

$$Y_i(t_2) = b_1 Y_i(t_1) + b_2 \phi + e_i \quad , \tag{1}$$

where $Y_i(t_2)$ = posttest score,

$Y_i(t_1)$ = pretest score,

ϕ = dummy variable (1 = treatment, 0 = control) indicating presence or absence of treatment, and

e_i = random error term that is assumed to be uncorrelated with treatment group membership.

The size of the regression weight b_2 in relation to its standard error leads to a judgment about the presence or absence of a statistically significant treatment. The sign of b_2 indicates the direction of the effect. In this particular example, a finding of a significant positive b_2 would be interpreted as evidence for the presence of a coaching effect. A finding of a negative b_2 suggests that coaching has a debilitating effect on posttest scores.

A more realistic model might be

$$Y_i(t_2) = b_1 Y_i(t_1) + b_2 \phi + d' z + e_i \quad , \tag{2}$$

where z = vector of variables which are assumed to be causes of self-selection, and d' = vector of regression weights associated with the self-selection causes. Relevant causes of self-selection might be negative self-perception of test-taking skills, motivation, ability to pay for formal coaching, etc.

In the FTC study we have no direct measures of these self-selection variables and thus must rely on proxies. To the extent that the pretest $Y(t_1)$ and the proxies only incompletely (or incorrectly) measure the causes of self-selection, the estimated regression weight, b_2, will be biased. In statistical parlance the assumption of the independence of the error (e_i) and the treatment no longer holds. Furthermore, if the coached group has an initially higher mean on the pretest and if the pretest and the other causes of self-selection which are either imperfectly measured or not measured at all are positively related to the treatment effect and the posttest score, then we can expect the ANCOVA model in (2) to overestimate the treatment effect. Conversely, if the missing or incompletely measured control variables are in general negatively related to the treatment effect and posttest score, then one can expect ANCOVA to underestimate the treatment effect. If they are uncorrelated with the treatment and the posttest (as would be the expectation in random assignment), then b_2 would give an unbiased estimate of the treatment effect. Obviously there are many other more complicated patterns of relationships which could occur and for which no simple predictions can be made about the direction of the bias.

Of particular concern here is the fact that motivation is either incompletely measured or not measured at all by the available demographic control variables. Since one of the consequences of motivation is learning performance, one would expect a measure of motivation to be positively correlated with both performance at the posttest and self-selection for the treatment (coaching). One might argue that differential motivation would be at least partially captured in the pretest score and/or other covariates in the vector z and thus be partially controlled, but in general that will not be true if the differential motivation is reflected (as is the likely case) in differential learning or growth rates. That is, even if differential motivation were completely measured, the statistical correction employed by ANCOVA is a static correction. The ANCOVA adjustment only makes a correction for differences found at the time when pretesting takes place. The treatment does not occur at the time of pretesting but at some later date when, if there is differential group growth, the groups will be further apart than they were at pretest time. The traditional ANCOVA correction adjusts for the differences found at pretest time and not the greater group difference that would occur (because of differential group growth rates) just previous to the intervention, which in this case takes place at an unknown period of time after the pretest.

This underadjustment bias in the presence of differential group growth can be made a little clearer if we examine the ANCOVA model in the following form:

$$\alpha = [\bar{Y}_p(t_2) - \bar{Y}_c(t_2)] - b\,[\bar{Y}_p(t_1) - \bar{Y}_c(t_1)] \quad , \qquad (3)$$

where $Y_p(t_2)$ and $Y_p(t_1)$ are posttest and pretest means for the coaching program group, $Y_c(t_2)$ and $Y_c(t_1)$ are the parallel measures for the control group, and α is the estimated treatment effect. The pooled within-group regression coefficient b is, of course, the adjustment index. If there is reason to believe that the following differential development or growth process is taking place:

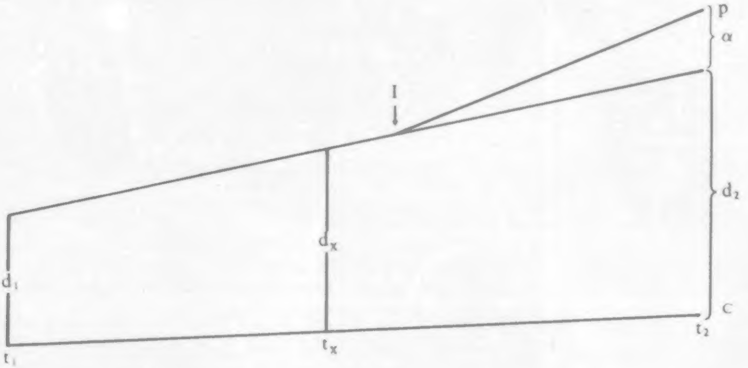


Figure 1

then any adjustment based on the initial group differences at pretest time (t_1) rather than at t_x will overestimate the treatment effect unless b, the adjustment index, is equal to d_2/d_1. In fact, any adjustment based on group differences at time t_x will still overestimate the treatment effect (although substantially less than an adjustment at t_1) unless the adjustment index is greater than 1.0. In general, most pre-post ANCOVAS based on data from nonequivalent control group designs yield bs $\cong$ 1.0. The "net effect" of the adjustment made by the FTC analysts using an ANCOVA model similar to that outlined in equation (2) was equivalent to using a b of approximately 1.0 in equation (3) for the verbal data and somewhat over 1.0 for the mathematics data Clearly, the traditional ANCOVA model as posed by the FTC analysts will not always be adequate if individuals are growing at a different rate over time. To deal adequately with data generated by growth processes, corrections for initial differences must take into consideration: (1) the time variable and (2) measures of differential group growth rate in the absence of formal intervention.

The FTC analysts did not investigate the possibility of differential group growth rates as a rival hypothesis. Their verbal ANCOVA model implicitly assumed (and its validity depended upon it) that the observed group differences at pretest time were invariant over time in the absence of a formal treatment. This assumption is reflected in their use of a "net" correction index of b $\cong$ 1.0. Conversely, since the FTC mathematics ANCOVA yielded a "net" correction index slightly over 1.0, the FTC mathematics results implicitly assumed a growth model. Campbell (1969) suggests that if two groups start out at time 1 with divergent means, those with the higher mean mature or grow at a greater rate than those with the lower mean. Campbell calls this the interaction of selection and maturation. The different groups are members of different populations living in different environments.* The different environments interact with differences in ability and create and maintain different levels of performance and different rates of growth. A special case of this differential growth phenomenon is the so-called "fan spread" model which postulates that increasing variability within group accompanies increasing mean differences over time (Kenny, 1975). The "fan spread" hypothesis is only one possible explanation and is sometimes used to arrive at a correction index when group growth data in the absence of intervention are not available. In its strictest sense a solution is obtained by the use of an overidentifying [illegible]triction depending on the stationarity over time of the ratio of mean differences to changes in the within-group variability. Rather than make any such assumptions, we will use the observed data to estimate group growth rates in the absence of intervention.

An interrupted time series design could provide an estimate of the relative growth rate in the absence of a formal intervention. That is, the following design uses growth data gathered between time t_0 and t_1 to

P	O	O	I	O
C	O	O		O
	t_0	t_1		t_2

---

*The FTC analysis of the group demographic variables does provide evidence that individuals in the self-selected population tend to come from different socioeconomic environments than the controls.

yield an estimate of both the program (P) and control (C) group rate of growth in the absence of a formal intervention. Fortunately, the majority of the individuals who attended the most successful coaching school (School A) had three testings with a coaching intervention taking place between time t_1 and t_2. Following the growth model development of Bryk and Weisberg (1977), we assume that the treatment effects a constant increment for each subject exposed to it. Denoting $G_i(t)$ as the actual growth for individual i under whatever conditions prevail, we then have at posttest

$$Y_i(t_2) = G_i(t_2) + \phi_i\alpha \quad , \tag{4}$$

where $\phi = 1$ if the subject is in the coaching program and $\phi = 0$ if the subject is in the control group. $G_i(t_2)$ is the cumulative growth of the *ith* subject at the time of posttesting. $G_i(t)$ is completely determined by two parameters: π_i and τ_i. That is:

$$G_i(t) = \pi_i(t - \tau_i); \; t \geqq \tau_i \quad , \tag{5}$$

where π_i is an individual's growth rate and τ_i is that point in an individual's development at which he achieves a prespecified although arbitrary level of performance. For the purpose of this study, the $E(\tau_i) = \tau$, that point in time when the two groups' growth lines intersect. At this point in time, the groups are at equivalent levels of knowledge. (See Figure 2).

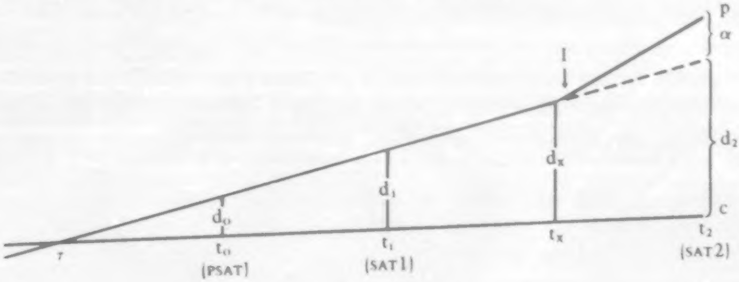


Figure 2

In this study, we are interested in estimating π for individuals in both the coaching program group and the control group based on the time interval between the first and second testing, i.e., before the intervention took place. Thus, the estimation of π for the *ith* subject in the coaching program is

$$\pi_{pi} = [Y_{pi}(t_{1i}) - Y_{pi}(t_{0i})]/(t_{1i} - t_{0i}) \tag{6}$$

That is, the growth rate is the difference between a subject's test scores at time t_0 and t_1 divided by the time interval measured in months. The expected value for an individual selected at random from the treatment group ($\mu_{\pi p}$) is $[\bar{Y}_p(t_1) - \bar{Y}_p(t_0)]/\bar{T}$, where $\bar{T}$ is the average time lapse between first and second testings for individuals in the coaching program group. Similarly $\mu_{\pi c}$ can be estimated for the control group from the data based on the first and second testing. It should be noted that in the case of random assignment $\mu_{\pi p} = \mu_{\pi c}$.

A more appropriate adjustment index which takes into consideration the possibility of differential group growth rates $\mu_{\pi p}$ and $\mu_{\pi c}$ as well as different Ts can now be derived. In terms of the general linear covariance model, which does not assume differential growth rates, we have the estimated treatment effect $\hat{a}$ in large samples as

$$\hat{a} \cong E[\bar{Y}_p(t_2)] - E[\bar{Y}_c(t_2)] - b_{\lim}\{E[\bar{Y}_p(t_1)] - E[\bar{Y}_c(t_1)]\} \quad . \tag{7}$$

Let a be an unbiased estimator of the treatment effect under the assumption of a differential group growth model; then

$$\begin{aligned} a = &\{E[\bar{Y}_p(t_2) - \mu_{\pi p}\bar{T}_{p2}] - E[\bar{Y}_c(t_2) - \mu_{\pi c}\bar{T}_{c2}]\} - \\ &b^*\{E[\bar{Y}_p(t_1) - \mu_{\pi p}\bar{T}_{p1}] - E[\bar{Y}_c(t_1) - \mu_{\pi c}\bar{T}_{c1}]\} \quad , \end{aligned} \tag{8}$$

where T_{p2} is the average time interval between τ and the third testing, t_2, for individuals in the coaching program. Similarly $\bar{T}_{p1}$ is the average time between τ and the second testing for the coached individuals. Parallel notation is used for the control group.

Then $\hat{a} - a = 0$ when

$$\hat{b}_{\lim} = b^* = \frac{\mu_{\pi p}(\bar{T}_{p2}) - \mu_{\pi c}(\bar{T}_{c2})}{\mu_{\pi p}(\bar{T}_{p1}) - \mu_{\pi c}(\bar{T}_{c1})} \tag{9}$$

Thus b^* represents the theoretically correct adjustment coefficient if indeed the differential growth rate model holds.

Figure 3 presents means and standard deviations of the verbal test scores at three points in time. Similar data is presented for mathematics test scores in Figure 4. The three testings consisted of the PSAT, the SAT taken for the first time as a junior, and the SAT

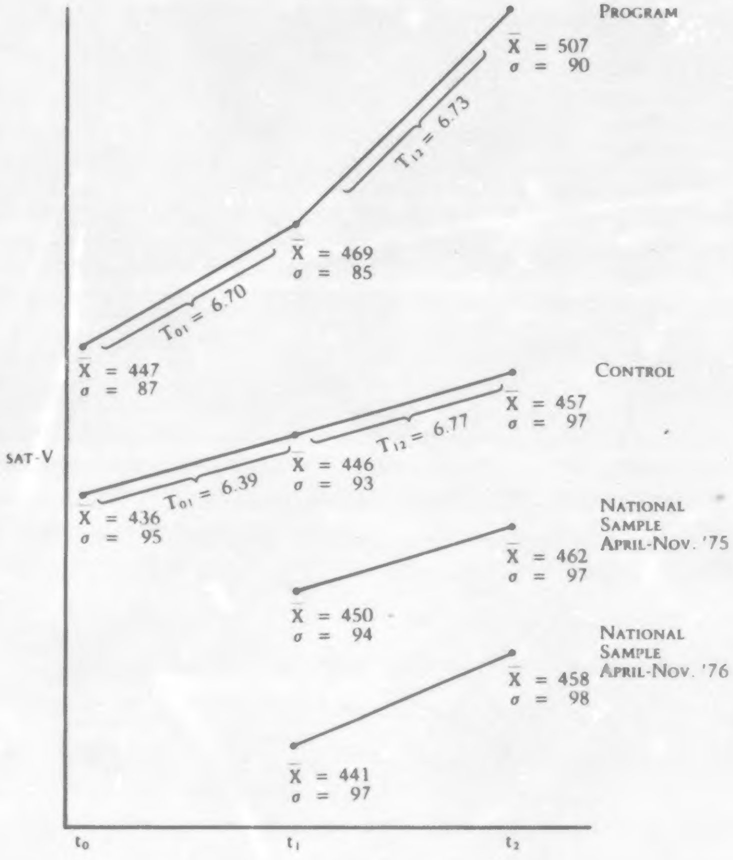


Figure 3

Verbal Gains Over Three Time Periods for Four Populations

taken for the second time as a senior. Only one coaching school (School A) had a large enough subject population to furnish sufficient subjects with all three data points. The coached sample included 192 individuals while the controls numbered 684.

The verbal data in Figure 3 suggests that the observed group differences at posttest may well be reflecting differential growth rates as well as treatment effects. The classic growth phenomenon of increasing means is demonstrated for both the coaching program people and the controls. While there appears to be a significant differential group growth rate in the verbal area, this does

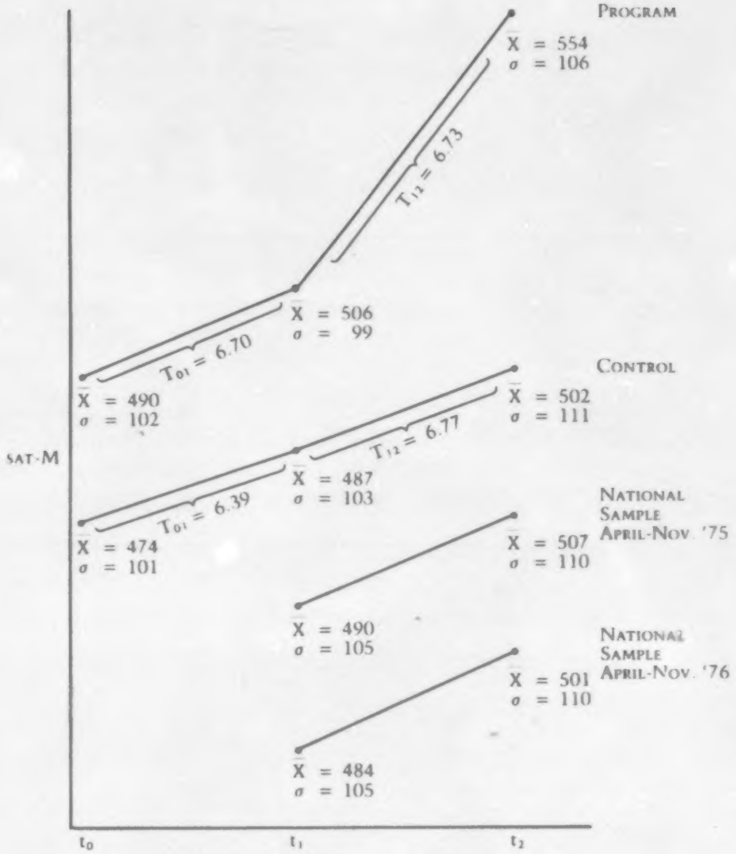


Figure 4

Math Gains Over Three Time Periods for Four Populations

not appear to be the case for mathematics. This seems reasonable since mathematics is a skill which requires practice if one intends to maintain or increase one's level of achievement.

Additional plots of the total College Board sample who were tested in April of their junior year and retested in November of the senior year are also shown. These plots are presented for 1975 and 1976 populations. Inspection of the verbal plots suggests that gains of 12-17 points from April to November appear to be commonplace for junior to senior retesters. The fact that the control sample shows a gain of 11 points over approximately the same interval suggests that they are reasonably representative of junior-senior retesters. The control group's rate of gain is consistent over both plotted time periods.

Table 1 presents ANCOVA and growth model adjustment indices and coaching effects, as well as the mean growth rates for the coaching program and control individuals for the verbal and mathematics areas. These estimates of growth are based on the time period between the first and second testing and thus are unconfounded with any formal intervention. Our estimates of growth rates assume linearity of growth but this would seem to be a reasonable assumption given the restricted time period and, even more importantly, given that the observed data for the control group conforms quite closely to a linear growth model. If one were to expect a deviation from linearity, it would more likely be in the direction of a steeper slope (in the absence of intervention) between the second and third testing since the motivational level would be at least as great during this time period as the previous time period.

The estimates of the coaching effect under the growth model are about 17 points on Verbal and 31 points on Math. This finding that the Verbal effect is substantially less than the Math effect is in marked contrast to the ANCOVA estimates of quite comparable V and M effects. The growth model appears to yield an estimate of the Verbal coaching effect that is more consistent with earlier studies and with expectations that Math, being generally more curriculum related than Verbal, might be more responsive to coaching or special preparation.

The Verbal results in Table 1 call into question the ability of the standard ANCOVA to correct for selection effects when they are present in the form of differential growth rates. The ANCOVA adjustment index (b) is approximately 1.0 for the Verbal data and somewhat paradoxically 1.25 for the Math data, which is the one situation which displays little or no evidence of differential group growth. However, since the growth rates appear to be static for the mathematics data, the ANCOVA model, which controls for all available measures of pre-existing differences, would appear to be the more defensible approach for estimating the mathematics "coaching" effect.

An additional point should be discussed here. The original FTC analyses as well as the reanalysis as outlined here do not take into consideration the possible biasing effect of using a fallible covariate (pretest) in adjusting for pre-existing differences. If the reliability of the tests increased with each testing, one would expect the two groups if truly from different populations to demonstrate increased divergence over time in the absence of intervention. This phenomenon would not seriously affect the growth model approach but very well could have a biasing effect on the ANCOVA estimate of the coaching effect. That is, ANCOVA would yield an overestimate of the coaching effect. Similarly, if the reliabilities are constant but less than 1.0 (as surely they must be), then ANCOVA would underadjust for pre-existing differences on the pretest, thereby yielding an overestimate of the coaching effect. The data were not available to investigate these possibilities.

Table 1

Mean Growth Rates, Adjustment Indices, and Estimated Coaching Effects Under Different Model Assumptions

	Average Growth Rate[1] *(Points/Month)*	*Adjustment Indices and Their Estimates of Coaching Effects*[1]			
		Growth Model		ANCOVA	
	μ_π	b_i^*	a_i^*	b_i	a_i
Verbal					
Program	3.294	1.485	16.946	.993	28.065
Control	1.654				
Mathematics					
Program	2.354	1.102	30.558	1.252	27.625
Control	2.048				

[1] The average gain in points per month is estimated in the absence of intervention. Growth model estimates b_i^* and a_i^* are the adjustment index and the estimated effect based on the group growth rate estimates in the first column. The ANCOVA estimates are the standard estimates arrived at using the list of control variables presented in Table 2, Appendix 3. The ANCOVA b_i is the net adjustment index, when used in the usual ANCOVA equation it yields the effect estimate a_i.

Table 2

Analysis of Covariance Models

	Dependent Variables			
	Verbal Posttest		*Math Posttest*	
Explanatory Variables	*b*	*S.E.*	*b*	*S.E.*
SAT 1-V	.8024	.0238	.1538	.0253
SAT 1-M	.1307	.0235	.7760	.0251
Rank	.1286	.1319	.3201	.1403
Sex	2.0659	3.5128	13.4507	3.7389
Latest Grade in English	.5944	.3134	-.5246	.3335
Latest Grade in Math	-.2336	.2559	.6165	.2723
High School Type	6.4684	3.6079	-8.5227	3.8402
Log (Income)	-1.8385	2.8870	-1.0852	3.0728
Coaching	28.0658	4.0403	27.6247	4.3004
Years of Math	-2.9292	2.2961	8.7692	2.4400
Years of English	7.0633	3.0307	-2.8102	3.2258
Multiple Correlation	R = .8803		R = .8964	

Summary and Conclusions

The FTC data based on individuals who attended the largest and most successful coaching school were reanalyzed. The examination of testings at three points in time suggested that: (1) the traditional ANCOVA approach used by the FTC was inadequate for the Verbal data because of self-selection effects which were at least partially captured in differential group growth rates, (2) more appropriate growth-related adjustment models yielded Verbal coaching effects about one-half the size reported by the FTC, (3) the Math data appeared to be somewhat more consistent with the ANCOVA model and thus the resulting FTC Math coaching-effect estimate is more likely to be reasonable given the available control variables. This is not to say that the FTC and the present estimates of the mathematics coaching effect are not overestimates (or underestimates), since the only self-selection causes that have been controlled for were those that were reflected in differential growth rates and/or available demographics. One missing piece of information is the reliability of the test scores at each administration. When individuals self-select to treatments, commonly held psychometric wisdom suggests that as the reliability of the pretest becomes significantly less than unity, one can expect the ANCOVA model to underadjust, thereby yielding an overestimate of coaching effect. In general, the extent of this source of error would be small compared to the consequences of using the standard ANCOVA model in an inappropriate situation.

If the reliability of the pretest was different for the two groups, then almost any ANCOVA or variation of ANCOVA would in general yield biased results. It would not be expected, however, that the group reliabilities would be sufficiently different to significantly change the above conclusions. An additional problem with any reliability correction is that while there is evidence here that the groups are from different populations and therefore regressing on different means, there is no way of knowing from the available data what the appropriate means are.

References

Bryk, A. S., & Weisberg, H. I. Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin*, 1977, **84**, 950-962.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.

Campbell, D. T. Reforms as experiments. *American Psychologist*, 1969, **24**, 409-429.

Federal Trade Commission, Boston Regional Office, *Staff memorandom of the Boston Regional Office of the Federal Trade Commission: The effects of coaching on standardized admission examinations.* Boston, Mass.: Federal Trade Commission, Boston Regional Office, September 1978.

Federal Trade Commission, Bureau of Consumer Protection. *Effects of coaching on standardized admission examinations: Revised statistical analyses of data gathered by Boston Regional Office of the Federal Trade Commission*. Washington, DC: Federal Trade Commission, Bureau of Consumer Protection, March 1979.

Kenny, D. A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin*, 1975, **82**, 345-362.

[Excerpt From Psychological Bulletin 1981]

Copyright 1981 by the American Psychological Association, Inc.
0033-2909/81/8902-0191$00.75

Time and Method in Coaching for the SAT

Samuel Messick and Ann Jungeblut
Educational Testing Service, Princeton, New Jersey

An exhaustive review of studies of coaching for the Scholastic Aptitude Test (SAT) revealed them to be methodologically flawed in a variety of ways. Despite these flaws and the attendant noise obscuring the results, some definite regularities emerged relating the size of score effects associated with coaching to the amount of student contact time entailed in the coaching programs. Rank-order correlations between these two variables were upwards of .7 for both SAT-V(Verbal subscale) and SAT-M(Math subscale). The relationship proved to be nonlinear, however, with arithmetically increasing amounts of score effect being associated with geometrically increasing amounts of student contact time, which in these data may also be serving as a proxy for increasing curriculum emphases on content knowledge and skill development as opposed to item review and practice. Thus, within the limitations of the available fragmentary data, there appear to be diminishing returns in SAT coaching effects, especially for SAT-V. According to the logarithmic models fit to the data, the student contact time required to achieve average score increases much greater than 20 to 30 points (on a 200- to 800-point scale) for both SAT-V and SAT-M rapidly approaches that of full-time schooling.

The controversy over whether or not the Scholastic Aptitude Test (SAT) is coachable is fueled by both misconception and miscommunication. The misconception centers mainly around the nature of the SAT, whereas the miscommunication centers mainly around the nature of coaching.

The SAT is not a measure of subject-matter attainment, such as the typical educational achievement test in biology or American history, nor is it a measure of innate intelligence or fixed endowment. The SAT measures developed abilities of verbal and mathematical reasoning and comprehension that are acquired gradually over many years of experience and use in both school and nonschool settings. By virtue of this gradual development, these intellective skills appear to be relatively difficult to improve markedly through brief courses of intervention in the final year or 2 of high school, when the SAT is typically taken. But since these abilities are learned, albeit in manifold ways through both education and experience, one would expect high-quality instruction over extended periods of time to improve them and hence to increase SAT scores. Thus, the question of the degree to which the SAT is coachable depends on what is meant by *coaching*, a point about which there has been not only marked contention but lax communication. Some writers restrict their usage of the term to practice on sample items and last-minute cramming (cf. Pike, 1978), whereas others include under the same rubric virtually full-time instruction at specialized preparatory schools for periods of 6 months or more (cf. Slack & Porter, 1980).

In an effort to uncover some regularities in the results of coaching studies that go beyond a simple scorecard of significant and nonsignificant findings or a conglomerate averaging of score effects from different coaching programs, we review all of the available studies of coaching for the SAT regardless of the way the term is interpreted. We view the issue as being much more complicated than the simplistic question of

---

We wish to thank Donald Alderman, Albert Beaton, John Carroll, Robert Linn, Frederic Lord, Donald Powers, Ledyard Tucker, and Bary Wingersky for their careful review of the manuscript and their many helpful comments. Special thanks go to John Tukey, whose suggestion that coaching effects be examined in terms of score gain per contact hour provided the impetus for this analysis.

Requests for reprints should be sent to Samuel Messick, Educational Testing Service, Princeton, New Jersey 08541.

whether or not coaching works for the SAT. Rather, the key questions are how much student time devoted to what kinds of coaching experiences yield what level of score improvements in comparison with the level of experiential growth that would have occurred anyway without those coaching experiences. This latter point, that coaching effects must be evaluated relative to the experiential growth in ability that may occur regardless of the coaching program, underscores the need for comparable control groups of uncoached students in studies of coaching effectiveness—but comparison with equivalent control groups has by no means been the rule. Since this point is critical in interpreting the import of experimental findings, the summaries of the various coaching studies draw special attention to the strengths and limitations of the study designs. Before these studies are examined in detail, however, further clarification of both the nature of the SAT and the nature of coaching is offered.

Coaching for Developed Abilities and Subject-Matter Learning

The SAT was developed as a measure of academic abilities to be used toward the end of secondary school as a standardized supplement to the high school record and letters of recommendation typically available to college admissions officers. The SAT was explicitly designed to differ from achievement tests in subject-matter fields in the sense that its content is drawn from a wide variety of substantive areas, not tied to a particular course of study, curriculum, or program. It assesses intellectual processes of comprehension and reasoning that are exercised to some degree in all subject-matter areas at all levels of schooling, as well as in response to real-life situations. The item content of the SAT attempts to sample the sort of generic cognitive skills underlying performance in college: reading with comprehension, understanding vocabulary, verbal reasoning, computational skills, quantitative reasoning, and problem solving.

There is a progression or ordering of educational tests ranging from measures of scholastic abilities at one pole, with content drawn from a variety of substantive areas, to measures of academic mastery at the other pole, with content specialized by subject-matter field (cf. Anastasi, 1980; Snow, 1980). The SAT falls toward the first extreme—it taps general cognitive processes that develop gradually over many years of experience and use both inside the classroom and in everyday life, and these processes should therefore be relatively difficult to enhance markedly through brief courses of intervention. The typical educational achievement test falls toward the other extreme—it taps specific knowledge and skills acquired through the normal course of classroom instruction or independent study, and such knowledge and skill should therefore be relatively responsive to instructional intervention, even in brief courses. Similarly, there is a progression of types of preparation for taking examinations ranging from simple practice on sample items at one extreme to intensive instruction aimed at developing ability and knowledge at the other extreme. What has come to be called *coaching* is here considered to fall anywhere in the broad range between these two extremes of practice and instruction, entailing some combination of test familiarization, drill-and-practice with feedback, training in strategies for specific item formats and for general test taking (including advice on pacing, guessing, and managing test anxiety), subject-matter review, and skill-development exercises. There is little question that the verbal and mathematical abilities measured by the SAT are learned, but there are large questions about how they can be taught. Thus the functional characteristics of any SAT coaching programs that prove to be effective would have important implications for educational practice.

The effect of coaching or special preparation programs on SAT performance has become a question of increasing concern to students, their parents, secondary schools, and colleges. A recent survey conducted in seven Northeastern states (Alderman & Powers, 1980) indicated that perhaps as many as one third of the secondary schools in the area offer a program of special preparation for SAT-V(Verbal subscale), with programs in SAT-M(Math subscale) probably at least as numerous. Commercial or proprietary coaching programs are also quite widely available in urban and suburban communities (Federal Trade Commission [FTC], 1978). Studies have been conducted periodically over the past 30 years to investigate the effects of both school-based and proprietary programs on student performance on the SAT. The results of these studies are examined with particular attention to problems of reseach design.

Problems of Method and Interpretation in SAT Coaching Studies

In the evaluation of coaching effectiveness, a key feature of research design is random assignment of examinees to coaching treatment groups and noncoaching control groups, for only with random assignment can we consider treatment effects to be independent of prior status on any of a host of personal or background characteristics. With random assignment, no systematic differences are expected between the experimental and control groups initially, and if effective control conditions are maintained, the only systematic difference that will eventuate is that one group will have received coaching and the other will not. In the absence of randomization, there is an inevitable equivocality in the interpretation of the results because some unmeasured personal characteristics might have influenced both the student's participation in the coaching program and that program's apparent effectiveness. That is, certain personal factors characteristic of students attending a particular coaching program, such as motivation or career aspirations, may be responsible, at least in part, for subsequent SAT performance that *appears* to be the result of the coaching experience. Thus the effects of selection or of self-selection are confounded with effects of the coaching treatment in nonrandomized studies. Consequently, selection factors afford plausible rival explanations for the results or for part of the results that might otherwise be identified as coaching effects. Selection factors include all systematic differences, whether measured or unmeasured, between the experimental and control groups that are correlated with the dependent variable, in this case SAT performance except, of course, that the experimental group received the coaching treatment, whereas the control group did not. If these systematic differences result from student choice of the treatment rather than from experimenter choice of the student, they are called self-selection factors.

In nonrandomized designs, researchers usually attempt to control statistically, using regression techniques like analysis of covariance, for those potential selection factors that have been measured. Attempts are also frequently made to analyze the data in alternate ways to assess the sensitivity or robustness of the findings under various plausible assumptions. However, there is no way to adjust statistically for selection factors that have not been assessed. Since very few of the studies reviewed here employ random assignment and those that do have other problems of maintaining realistic control conditions, the specter of selection bias arises continually as a plausible altenative interpretation of apparent coaching effects.

As our review underscores, the available studies of coaching for the SAT are methodologically flawed in various and divergent ways. Most are subject to the problem of selection bias just discussed, which severely compromises interpretations of the source or determinants of score effects. Others were subject to disturbances of student motivation or of control conditions, which very likely introduced biases in estimating the size of score effects. Still others were based on small samples of coached students, which resulted in estimated score effects that were imprecisely bounded. Several studies suffered from combinations of two or more of these problems. In searching for regularities among the findings, especially between the size of score effects and key characteristics of the coaching programs, we attempt to discount studies having likely biases in their score estimates while at the same time trying to take into account the greater variability of estimates from studies with small sample sizes. Although the statistical significance of results is documented (see Table 1), the issue of coaching is addressed more in terms of estimation of score effects than of rejection of the null hypothesis. Finally, we turn to the problem of interpreting the determinants or sources of the educed relationships and confront once again the pervasive confounding influence of selection bias.

Although there was considerable variation in the score increases observed for particular groups of students and for particular coaching programs, the studies conducted in the 1950s and 1960s as well as two more recent studies (Alderman & Powers, 1980; Evans & Pike, 1973) yielded average score increases relative to control groups of about 9 points on SAT-V and about 13 points on SAT-M, on a score scale ranging from 200 to 800 points with a standard deviation of about 100 points (Messick, 1980). If the average effects for two commercial coaching schools studied by the FTC (1979; see also Stroud, 1980) are added, the average SAT-V increase becomes 14 points and the average SAT-M increase becomes 15 points. Studies lacking any control groups whatsoever yielded larger adjusted score effects—on the average, 38 points on SAT-V and 54 points on SAT-M—but since they differed from control group studies not only in design characteristics but in critical program characteristics as well, their interpretation is especially problematic.

Studies With No Control Groups

Three studies—those by Pallone in 1961, by Marron in 1965, and by Coffman and Parry in 1967—lacked any control group for evaluating unusual patterns of score change. Pallone examined the effects of short- and long-term intensive developmental reading courses on SAT scores of students in a private school for boys. The courses were undertaken "for students in their final year of precollege work, including a large number of high school graduates who were completing a year of post-high school study in preparation for entrance into the U.S. government academies" (p. 655). According to Pallone, to improve the skills measured by the SAT, "not 'coaching' methods, but instruction of a developmental nature in reading and vocabulary skills was indicated. Improvement in scores could be expected only if the basic skills measured by the test were first strengthened" (pp. 654–655). This program provided focused instruction to strengthen reading achievement along with intensive practice in reading skills, including such special skills as skimming and critical reading, as well as a brief analysis of typical verbal analogy test items. Approximately 20 students participated in a 6-week summer pilot program that met for 90 minutes daily. An average score increase of 98 points was obtained on SAT-V from the March to August SAT administrations. The long-term program covered a 6-month period with daily meetings of 50 minutes each. For some 80 students who completed this long-term course, an average SAT-V increase of 109 points was reported over a 12-month period from pretest to posttest, although the difference in mean scores in Pallone's Table 3 (p. 656) is only 84. The students who participated in the summer program also completed the long-term reading course, and their mean increase in Verbal score over 12 months was almost 122 points, or an average of about 24 points over the increase reported after the summer course. The special quality of the sample and the lack of control groups severely limit the implications of these findings vis-à-vis coaching. Furthermore, the instructional focus on skill development and the intensive and long-term nature of the programs put Pallone's efforts close to what ordinarily would be considered "instruction," in contradistinction to "coaching," as Pallone himself insisted.

In the absence of control groups of similar students at this preparatory school who were not taking Pallone's course, it is difficult to assess the import of these score gains. Pallone (1961) suggested comparing them with normal expectations of gains of about 35 points on SAT-V during the final secondary school year, which would yield an instructional or program effect of about 83 points for the summer pilot program (prorated for the 5 months between pre- and posttests) and about 74 points for the long-term program (which had 12 months between pre- and posttests). Or, if the difference in mean scores is taken as 84 as in Pallone's Table 3, the long-term instructional effect is about 49 points. This is not a very satisfactory comparison, however, since Pallone's students were not representative of students in their final year of high school who take the SAT. In a recent review article on the SAT, Slack and Porter (1980) suggested comparing Pallone's results with average gains in national administrations of junior- to senior-year retesters having the same initial average score levels as Pallone's students, which yields prorated instructional effects of 85 and 79 (or 54) points, respectively, for the summer and long-term programs. Again, this is not a very satisfactory comparison because Pallone's private school students were not a representative sample of the national population of test repeaters. Pike (1978) suggested comparing Pallone's results with average gains of control students in superior schools from other studies of proprietary programs, and this comparison leads to instructional effects of about 75 and 53 (or 28) points, respectively, for the summer and long-term programs. If compared with score gains of control students in other coaching studies who had average initial score levels roughly comparable to Pallone's groups, the adjusted instructional effects are 80 and 65 (or 40) points, respectively.

The point is that in the absence of comparable control groups, no generally satisfactory estimate of instructional effects can be obtained. If an average is taken across the four adjustments just suggested, the resulting estimates of instructional effects are 81 and 68 (or 43) points, respectively, for the summer and long-term programs. These values are very likely still overestimates, however, because none of the comparisons can take into account the highly self-selected nature of Pallone's (1961) private school students, many of whom were completing a post-high-school year particularly motivated to increase their chances of entering service academies or selective colleges. Nevertheless, given the overall size of the effects, even with somewhat larger adjustments, it seems likely that Pallone's intensive summer and long-term efforts at "instruction of a developmental nature" succeeded in strengthening basic skills measured by the SAT.

Marron (1965) examined SAT score gains for students at 10 well-known preparatory schools that specialized in preparing high school graduates for admission to the service academies and selective colleges. The instructional programs entailed "six months of full-time exposure to course content that is directly related to the verbal and mathematics College Board tests (both Aptitude and Achievement)" (p.1). A special administration of SAT-V and SAT-M at all 10 schools served as the pretest, whereas the posttest was a regular SAT administration 6 months later. It should be noted that if the level of motivation and effort on a special pretest that did not count for college admission was not comparable to that on the regular posttest, the instructional effects in Marron's study would likely be overestimated. This issue did not arise for Pallone (1961) because both pre- and posttests in his study were regular SAT administrations.

Since significant differences were obtained among the 10 schools with respect to both the pretest scores and the posttest scores and these latter differences remained significant in analyses of covariance adjusting for pretest levels, the overall results were reported separately for groups of schools having nonsignificant differences within group. Score gains on SAT-V were 77 points for Group 1 (2 schools, $N = 83$), 56 points for Group 2 (6 schools, $N = 600$), 47 points for Group 3 (1 school, $N = 5$), and 35 points for Group 4 (1 school, $N = 26$); the weighted average SAT-V increase over all groups was 58 points. Score increases for SAT-M were 83 points for Group 1 (4 schools, $N = 232$), 78 points for Group 2 (3 schools, $N = 405$), and 72 points for Group 3 (3 schools, $N = 78$); the weighted average SAT-M increase over all groups was 79 points. Again, in the absence of control groups, it is difficult to appraise the size of these instructional effects.

Marron (1965) suggested comparing the SAT score gains with the gains considered typical for males in their senior year in secondary school, which he reported based on College Board data as 40 points for SAT-V and 43 points for SAT-M over a 10-month testing interval. Prorated for the 6-month testing interval in Marron's study, this yields an adjusted weighted average of 34 points for SAT-V and 53 points for SAT-M. As was done for Pallone's (1961) results, the suggested adjustments of Slack and Porter (1980) and Pike (1978) were also applied to Marron's figures, along with an adjustment based on score gains of control students in other coaching studies who had average initial score levels roughly comparable to the scores of Marron's groups. Taking the average of all four of these adjustments, the resulting weighted average values are 35 points for SAT-V and 54 points for SAT-M. But again, none of these adjustments is very satisfactory because none of the suggested comparisons takes into account that Marron's students were highly self-selected, thereby leaving important factors of differential motivation and growth uncontrolled. In any event, the relevance of Marron's study to the issue of coaching is arguable, since 6 months of full-time exposure to course content directed at verbal and mathematical knowledge and skills would ordinarily be considered instruction.

In an attempt to explore further the effects of developmental reading instruction on SAT-V scores, especially in light of Pallone's (1961) findings, Coffman and Parry (1967) studied three groups of college freshmen who took the SAT-V before and after completing a course in accelerated reading. The course was described as stressing speed of reading with relative accuracy. Pre- and posttest scores based on special administrations of the SAT were available for two small groups of 10 and 9 students who elected to take an 8-week course that met 6 hours each week. Pre- and posttest SAT-V scores were also available for 25 students whose course met 3 hours a week for 15 weeks. For students in the 8-week course, SAT-V scores increased 3.5 points in one group and 9.9 in the other. A 28.9 mean loss was observed for the group taking the 15-week course, possibly because of problems of test administration and score equating, since time constraints dictated the use of a shortened SAT-V for that group. It may also reflect problems in the motivation of students who take a special SAT when they are already in college, as may the relatively modest score increases in the other two groups—although all of these students were presumably motivated to enroll in the course, which explicitly entailed taking the SAT. Again, the lack of control groups seriously impairs the usefulness of these results.

Studies With Nonequivalent Control Groups

A second methodological weakness obtains in four other coaching studies: Although each incorporated control groups, the control students attended different schools from those providing special preparation for experimental students or else were drawn from other extrinsic sources such as test-score files, thus confounding coaching effects with school effects and with myriad other self-selection factors.

Dyer (1953) studied seniors at two highly selective independent schools for boys—225 students at one school served as the treatment group, and 193 students at the other school served as the control group. The students at both schools took a special pretest SAT and, 6 months later, a regular SAT, which served as the posttest. The experimental group completed 12 verbal practice exercises in 30- to 60-minute sessions and 5 math practice exercises in 60- to 90-minute sessions. The control variables in an analysis of covariance were initial SAT scores, number of years each student had been enrolled in school, and the number and level of foreign language and mathematics courses taken in the senior year. The estimated increase in score for the treatment group over the control group was 4.6 points in Verbal and 12.9 points in Math. When the students were divided into those who were not taking mathematics courses as seniors and those who were, the no-math boys who were coached gained over 29 points more than those who were not coached. In contrast, it was found that the boys taking mathematics who were coached gained 3.3 points more than those taking mathematics who were not coached. These findings provide the first example in studies of coaching of an interaction between size of effects and student background characteristics, alerting us to the more general possibility that coaching programs, like other forms of teaching, may have differential effects for different kinds of students.

French (1955) conducted a coaching study employing dual treatment and control groups at three schools: The 158 students at School A pursued their regular courses with no attempt at special preparation for the SAT and served as a control group for coaching in both Verbal and Math; the 110 students at School B served as a treatment group for Verbal and as a control group for Math; the 161 students at School C served as a treatment group for both Verbal and Math. The special preparation program in Verbal at School B differed from the one at School C in that the former primarily emphasized vocabulary for a total of 4½ hours, whereas the latter reviewed 10 verbal exercises more representative of SAT-V skills. The pretest was a special administration of the SAT, with a regular administration 6 months later serving as a posttest. An analysis of covariance, using pretest scores as the control variable and posttest scores as the dependent or outcome variable, showed an advantage in Verbal score for the coached groups of 18 points at School C and 5 points at School B (the one with the vocabulary coaching program). The score increase in Math was 6 points when compared with scores of School A and 18 points when compared with scores of School B. When broken down by sex and current enrollment in math courses, the SAT-M data indicated that coached boys not studying math at the time of coaching showed greater increases over control students than did coached boys who were studying math, by about 4 points in one school and 10 in the other. This pattern of higher coaching effects for boys not currently studying math is consistent with the Dyer results. In contrast, coached girls not studying math at the time showed *smaller* increases over control students than did coached girls who were studying math. Coached girls currently studying math exhibited score increases of about 20 and 30 points over the noncoached girls currently studying math in the two control schools, whereas the coached girls not studying math exceeded their noncoached counterparts in the two math control schools by only 1 and 4 points. The interaction between size of coaching effects and current enrollment in math courses thus appears in these data to be moderated by the student's sex.

Dear (1958) undertook a study to determine whether longer periods of coaching in small groups—two class periods a week for 6 weeks and for 12 weeks—were likely to be more effective than the shorter term coaching studied by Dyer (1953) and by French (1955). In Dear's study, six public and four private secondary schools were chosen randomly from a list of schools in which at least 15 students had taken the May SAT as juniors. A treatment group from each school was selected at random from students who volunteered for coaching. A second group of nine schools drawn at random from the same geographical region—the New York-New Jersey-Greater Philadelphia area—served as control schools. Three students were selected from each school from each of three ability levels—90 coached students and 81 control students. Of these, 71 coached and 79 uncoached students took the SAT at regular administrations in May and again ten months later in March. The coaching program began in mid-November and continued through mid-March, with weekly coaching sessions supplemented by 1 additional hour of homework each week. Most students repeated the SAT in January (8 months after the pretest), which was halfway through the coaching period, and again in March (10 months after the pretest). The January results showed about a 22-point advantage for 60 coached students on Math, but a 2.5-point disadvantage on Verbal (relative to the average adjusted score increase of the control group). The March advantage for the coached students on Math was about 24 points. The Verbal score results, unfortunately, were not determined because of a significant difference in the slopes of the regression lines for the coached and uncoached groups. In contrast to Dyer's (1953) results, the greatest coaching gains on SAT-M were for students currently studying math, but no comparison can be made with French's (1955) finding of a sex-moderated interaction, since the Dear sample was not broken down for separate analyses by sex of student.

An ambitious investigation of the effects on student test performance of commercial coaching for the SAT was undertaken by the FTC in 1978, with extensively revised statistical analyses being issued the following year (FTC, 1979). Students enrolled in two New York City area commercial coaching schools during the 3 testing years of 1974 to 1977 served as the experimental or treatment group, and a random sample from College Board test-score files of uncoached persons who took the SAT during the same 3-year period in the same greater New York metropolitan area served as a control group. Six subgroups were examined: (a) high school juniors taking the SAT for the first time in April 1975 (76 coached and 607 uncoached students); (b) juniors taking the SAT for the first time in April 1976 (247 coached, 617 uncoached); (c) seniors taking the SAT for the second time in November 1975 (98 coached, 396 uncoached); (d) seniors taking the SAT for the second time in November 1976 (177 coached, 387 uncoached); (e) all high school students taking the SAT for the first time on any test date during the 3-year period (417 coached, 1,763 uncoached); and (f) all high school students taking the SAT for the second time during this period (316 coached, 1,267 uncoached). Statistical analyses were actually based on smaller samples than these largely because of missing student descriptive data.

When demographic and personal characteristics of the experimental and control groups were contrasted, it was found that the coached group was significantly higher than the uncoached group in high school class rank, parental income, most recent English grades, most recent math grades, and number of years of math taken. In addition, the coached group included significantly more nonpublic school students than the uncoached group. Although statistical adjustments were made for these and other preexisting group differences for which measures were available, there was no way to take into account unmeasured factors also likely to differentiate the groups, such as motivation or level of parental education. In the absence of random assignment of students to coached and uncoached groups and especially in view of the large and extensive differences confounded in the nonrandomized groups ultimately analyzed, any score effects derived from these data must be interpreted as combined coaching/self-selection effects.

The FTC used multiple regression techniques to control for group differences on the Preliminary Scholastic Aptitude Test (PSAT)—or on the first SAT when two were taken—as well as for differences on several relevant background variables. Although they found negligible effects for students at one commercial coaching school, statistically significant effects were obtained for students at the other school, where the impact for SAT-Verbal was estimated to be 30 and 27 points, respectively, for first- and second-time SAT takers over the pooled time periods. Score effects of 19 and 28 points, respectively, were estimated for SAT-Math over the same periods. The program at the school in which students exhibited significant average score increases (School A) involved 10 4-hour sessions for a total of 40 hours of coaching, whereas that at the school showing negligible overall effects (School B) involved a total of 24 hours of coaching. Again, because of uncontrolled differences between the coached and uncoached groups, these estimated score increases of 20 to 30 points for both SAT-V and SAT-M at one coaching school represent combined coaching/self-selection effects.

Two major reanalyses of the FTC (1978, 1979) data were subsequently undertaken to address a broader set of issues with more powerful analytical techniques. A third reanalysis (National Education Association, 1980) reexamined some of the same ground covered by the FTC analysis but in a relatively primitive fashion, adding neither novel findings nor better controlled or more precise estimates. In the first major reanalysis, Stroud (1980) applied an analysis of covariance type of model that differed from the FTC analysis in two important respects: The FTC analysis used a pooled regression equation across both coached and uncoached students using, along with other variables, verbal pretests and background measures (such as PSAT-V and English grades) in predicting score effects on SAT-V but quantitative pretests and background variables (such as PSAT-M and math grades) in predicting SAT-M effects. In contrast, Stroud employed a regression equation defined by the performance of uncoached students only and, in addition, included both verbal and quantitative pretests and background variables in predicting coaching effects for each area on the SAT. This analysis is not only more precisely controlled through the inclusion of additional covariates, but it more appropriately contrasts the performance of coached students with predicted score levels they would have expected had they been uncoached students with their same values on predictor variables. This approach results in valid estimates with fewer assumptions and facilitates straightforward examination of interactions (Cochran, 1968).

Stroud's (1980) reanalysis indicates that, given their background characteristics and pretest levels, students enrolled in one of the two coaching schools studied by the FTC (School A) obtained significantly higher SAT scores than did uncoached students by about 20 to 35 points in both Verbal and Math—the same neighborhood as the FTC estimates. The estimated effects for students at the other school (School B) averaged about 5 points in Verbal and 7 in Math and were not statistically significant. These are estimated combined effects due to coaching and self-selection, since it is not possible, as noted previously, to estimate coaching and self-selection effects separately with these data. Stroud also investigated the possibility of interactions between size of effects and the student background variables included as covariates. No interactions were uncovered at Coaching School A for either SAT-V or SAT-M. At Coaching School B, however, statistically significant and independent interactive effects were obtained on SAT-V for race and self-reported parental income. On the average, even though their number was quite small ($N = 13$), black students at School B exhibited significantly larger coaching/self-selection effects on SAT-V than nonblacks, and students reporting low family income exhibited significantly larger coaching/self-selection Verbal effects than those reporting high family income (Messick, 1980, pp. 46–51). In contrast to the findings of Dyer (1953) and French (1955), no differential score effects were uncovered for SAT-M, either as a function of years of math taken or of sex or of any of the other covariates.

In the second major reanalysis of the FTC data, Rock (1980) employed a statistical model that takes account of differential rates of growth in SAT scores over time, if they occur, for the coached and noncoached groups. If growth is occurring in the dependent variable (i.e., in the verbal and mathematical abilities measured by the SAT), then one of the key ways in which nonrandomized treatment and control groups might differ prior to any coaching intervention is in the rate of this intellectual growth. The statistical growth model employed by Rock does not adjust for self-selection factors or other group differences on background variables unrelated to differential growth, as analysis of covariance can. However, it does correct for those self-selection effects embodied in differential group growth that were not predicted from the available covariates used (Bryk & Weisberg, 1977). Since direct application of the model requires estimation of differential growth rates from time series data, even if only linearly from two testing occasions prior to coaching, only students having three test scores could be considered. For the treatment or experimental group, Rock included only those students at the largest coaching school (School A) for whom three test scores were available, a PSAT and two administrations of the SAT. The control group included only those uncoached students for whom these same three test scores were also available. The treatment group numbered 192 students and the control group, 684.

The coached students performed better than the uncoached students on the PSAT and were higher as well in high school rank-in-class and family income. As would be expected, the coached students also scored higher than their uncoached cohorts on the first SAT administration—but in the case of the Verbal area scored differentially higher. That is, during the period from taking the PSAT to taking the initial SAT, prior to attendance at the coaching school, the verbal skills of the students who subsequently were to be coached appeared to grow more rapidly than those of the uncoached students. In Math, however, both the coached and uncoached students showed similar group growth rates during this preintervention period. When the confounding effects of differential group growth rates were controlled for, the estimated coaching/self-selection effect for SAT-V dropped to about 17 points, whereas that for SAT-M, by virtue of not exhibiting differential group growth rates in these data, remained at about 30 points. Thus, by taking differential group growth rates into account, the estimated SAT-V effect was reduced to about half that in the FTC (1979) report or the Stroud (1980) reanalysis. Furthermore, the fact that Rock's estimate of the Verbal effect is substantially less than the Math effect is consistent with earlier findings and with expectations that Math, being generally more curriculum related than Verbal, might be more responsive to coaching or special preparation.

Studies With Matched Control Groups

A third methodological problem occurs in studies in which control and experimental students, although from the same school, are not assigned randomly but are matched on selected measures, thereby still permitting systematic differences between the groups on other nonassessed variables.

Frankel (1960), in the first study ever published on the effects of commercial coaching on SAT scores, selected 45 high school students who had taken commercial coaching courses and matched them with 45 control students from the same high school on the basis of (a) having taken the same two regular SAT administrations to serve as pretest and posttest, (b) having pretest scores within 10 points of each other on both Verbal and Math, and (c) being of the same sex. Within each pair, one student had taken a commercial coaching course involving roughly 30 hours of coaching in classes of about 25 students during the 7- to 8-month period between pre- and posttests. In brief, Frankel found an 8.4 point advantage for coached students on Verbal and a 9.4 point advantage on Math.

Whitla (1962) compared the score increases of 52 students who had attended a 10-hour course in improved study habits, reading skills, and math concepts at a proprietary school in Boston with the score increases of a comparable group of 52 students from the same area who had not taken a coaching course. All of the students had taken a regular SAT in March or May of their junior year and, in addition, took a special SAT in the fall of their senior year when the study began; the January regular SAT administration provided the posttest. The average Verbal and Math scores of the two groups were within 1 point of each other on the spring SAT and within 2 points on the fall pretest, suggesting that the two groups were not only well matched in terms of initial level but also in terms of growth rate over this period. Whitla found an 11-point advantage for the coached group on Verbal but a 5-point disadvantage on Math when the posttest was compared with the fall pretest; there was a 10-point advantage on Verbal and a 7-point disadvantage on Math when compared with the spring SAT. Note that the term *disadvantage* does not signify a score decrease on the part of the coached group, but rather a failure of their scores to increase as much as those of the control group.

Studies With Randomized Control Groups

Three coaching studies have used a randomized design. The first one, by Roberts and Oppenheim (1966), utilized the PSAT as both pretest and posttest. In contrast to earlier coaching studies that involved highly selective and effective private schools and specialized or suburban public schools, the Roberts and Oppenheim study was undertaken to investigate whether students receiving less adequate instruction might especially benefit from special preparation. Data were collected from 18 predominantly, if not entirely, black secondary schools in rural and urban Tennessee. In 6 schools coaching consisted of special instruction in verbal material, in 8 schools coaching was for mathematics, and in 4 schools no special instruction was provided. Within the treatment schools students were assigned randomly to coached and uncoached groups. The instruction was provided in 15 half-hour sessions over a 4- to 6-week period. The results showed small increases for the coached groups over the control groups: about 1½ (1.44) points on PSAT-V and less than 1 (.81) point on PSAT-M, increases that correspond to about 14 points on SAT-V and 8 points on SAT-M. This advantage of coached over control groups, however, was due as much to score decreases on the part of the control students, possibly signaling problems in motivation or attrition, as to score increases on the part of the coached students.

A second randomized study, conducted by Evans and Pike (1973; Pike & Evans, 1972), examined intensive coaching efforts in the math area. A sample of 509 students in 12 schools participated in the study. The coached students received 21 hours of instruction and 21 hours of homework, over a 7-week period, directed at one of the following item types: Regular Math (RM), Data Sufficiency (DS), or Quantitative Comparisons (QC). Three randomly chosen groups of students were defined in each school: one to be instructed in QC, one in either RM or DS, and one as a control group. These groups took the SAT first in October, which served as a pretest, and again in December (posttest) and the following April (delayed posttest). The pretest and posttest were special administrations of the SAT, whereas the delayed posttest was a regular administration. The three experimental groups were given special preparation during November and December, and the control group received instruction after the December posttest and prior to the April delayed posttest. On this schedule, all groups received instruction in test-taking skills—becoming familiar with test directions, pacing, and appropriate strategies for guessing and using partial information. All groups also had instruction in math content—numerical facts, numerical and basic algebraic skills, and in particular mathematical areas such as inequalities. In addition, there was practice on one of the item types for students in each respective experimental group. The study revealed score increases beyond those experienced by the control group for each of the three experimental groups coached on a particular item type—11 SAT-M score points for QC, 19 points for DS, and 25 for RM. However, because the Evans and Pike (1973) study was designed to investigate the relative susceptibility of three item formats to special instruction, it is difficult to say just how large the effects would be in terms of coaching for SAT-M. Their best estimate of score increases reflective of coaching for all four groups over the total period from October to April was about 25 points. The average increase over the control group for the three experimental item-type groups, weighted according to their respective sample sizes, was 16.5 points over the period from the October pretest to the December posttest. Pike (1978) later conjectured that, still keeping within the total 21 hours of special preparation, a judicious combination of instruction for both RM and DS, the two major item types then in SAT-M, would be expected to yield coaching or special preparation effects of about 33 points.

In the third randomized study, Alderman and Powers (1980) investigated the effectiveness of existing secondary school programs that had been initiated by the schools to improve the performance of students on the SAT-V scale. Students at each of eight schools for whom PSAT scores were available were randomly assigned to a special preparation group or to a control group. Access to the same preparation course was delayed for control students for the purpose of this study. A special administration of a retired form of the SAT was used as the posttest. Across the eight schools the overall increase in SAT-V attributable to special preparation was about 8 points, which is statistically significant at the .05 level. The actual effects ranged from −3 points at one school to 28 at another. Differences in effectiveness among the coaching programs were not statistically significant from school to school, however, and the best statistical estimate of the range was from 4 to 16 points. Nor, apparently, did the control groups react in comparable fashion from school to school, possibly reflecting differences in motivation or seriousness in approaching a special SAT that did not count for college admissions. The largest school effect of 28 points, for example, resulted in part from a control-group decrease almost equal in magnitude to the treatment-group increase (−11 score points in going from the converted PSAT to the SAT for the control group versus 13 score points for the treatment group, which yields a 28-point school effect when covariance adjustments are made). An attenuated form of this pattern occurred at two other schools, whereas the remaining five schools showed varying degrees of score increase for both treatment and control groups.

Another, more subtle methodological problem has emerged in the process of reviewing these coaching studies having randomized control groups, namely, the problem of engendering and maintaining realistic motivation for taking the posttest SAT, especially for uncoached control students. Developing realistic motivation and effort for taking pre- and posttests is a common requirement of all coaching studies. However, the three experiments that employed randomized control groups also used as the posttest a special administration of the SAT or PSAT rather than a regular administration, so that the control students would not have scores counting on their record before having access to special preparation. But by virtue of not counting for college admissions, these special administrations may have been viewed to some degree as practice tests, thereby eliciting less motivation and effort than would a regular SAT administration. As in the study by Roberts and Oppenheim (1966), warning signals suggesting this possibility were noted in the Alderman and Powers (1980) study. As discussed earlier, control groups in three schools were found to exhibit score *decreases* in going from a regular PSAT to a special SAT, even though the expectation is for a score *increase* from an October PSAT to a spring SAT of upwards of 10 or 12 points—this is the basal estimate provided by national administration samples. Such control-group score decreases substantially complicate the interpretation of intercept differences as estimates of coaching effects (Messick, 1980, pp. 20–22).

As noted, control-group scores also decreased in the Roberts and Oppenheim (1966) study, which similarly employed a special administration posttest, in that case a special PSAT. Unlike Alderman and Powers (1980), who used a regular administration of the PSAT as a pretest, the Roberts and Oppenheim pretest PSAT was a special administration as well. None of the nonrandomized coaching studies reviewed earlier displayed control-group score decreases, and the posttest in all of those studies was a regular administration of the SAT—although Dyer (1953), French (1955), and Whitla (1962) did use special pretest administrations and thereby introduced the possibility of other biases. The remaining randomized study, conducted by Evans and Pike (1973), employed special administration SATs as both pretest and posttest, but their delayed posttest was a regular SAT administration. Pike's (1978) subsequent interpretation of score increases from posttest to delayed posttest as reflecting the long-term consolidation or continuance of gains due to coaching becomes jeopardized from this vantage point by the plausible rival interpretation that those score gains instead reflect increases in motivation and effort in going from a special administration to a regular one.

Comparison of Results Across Studies of Coaching

For numerous reasons, including the diversity of design limitations and the differences in sample sizes, it is difficult to compare results across these several studies in a meaningful way (cf. Pike, 1978). Table 1 represents one such attempt for those studies having some type of control group. The size of coaching effects reported there were calculated uniformly as follows: When analysis of covariance was performed, the reported values are intercept differences between the experimental and control regression lines, weighted in the case of multiple experimental or control groups by their respective sample sizes. In four studies not reporting analyses of covariance, the values in Table 1 are average score increases of experimental over control groups, again weighted in the case of multiple experimental or control groups by their respective sample sizes. Two of these latter studies (by Frankel, 1960, and by Whitla, 1962) involved statistical matching, and two (by Roberts & Oppenheim, 1966, and by Evans & Pike, 1973) involved randomization. Averaging these results over all of the studies in Table 1, weighting in each case by the size of the experimental sample, yields 14.3 points for Verbal and 15.1 points for Math (the unweighted averages are 10.4 for Verbal and 13.0 for Math).

For those studies having no control groups, a summary is provided in Table 2. The special preparation programs summarized there focus on verbal or mathematical content knowledge and skill development and entail the largest amounts of student contact time of any of the studies reviewed, which would ordinarily lead one to characterize them as instruction rather than coaching. These programs range from 45 hours of student contact time over 6 weeks to 48 hours over 8 weeks to roughly 100 hours over 6 months to virtually full-time over 6 months or approximately 600 hours (assuming six 50-minute periods a day for 24 weeks). The 45-, 48-, and 100-hour programs were devoted solely to special preparation in Verbal, whereas the full-time programs involved both Verbal and Math preparation, presumably with each receiving roughly equal 300-hour coverage. In contrast, the most intensive of the control-group studies summarized in Table 1 were 40 hours of student contact time divided, presumably equally, between Verbal and Math (Stroud, 1980) and 21 hours over 7 weeks devoted to special preparation in Math only (Evans & Pike, 1973); in addition, School H in Alderman and Powers (1980) entailed 45 hours of student contact time over 10 weeks in Verbal preparation only.

In the absence of control groups, instructional or program effects were estimated in Table 2 in the manner described earlier—that is, by adjusting the average score gains reported in each study by the average of four adjustments, those suggested by (a) the authors of the original articles, (b) Slack & Porter (1980), and (c) Pike (1978), as well as (d) the average score gains of control students in other coaching studies who had roughly comparable initial score levels. Averaging these estimates over all the studies in Table 2, weighting in each case according to group size, yields 38 points for SAT-V and 54 points for SAT-M (the unweighted averages are 39 points for Verbal and 53 points for Math). Given that the adjustments were dubious and provisional and that the students in each program were highly self-selected, these values are still probably overestimates of program effects. But their general magnitude suggests that the verbal and mathematical reasoning skills measured by the SAT may be enhanced to a measurable degree by long-term and intensive instruction, at least for highly motivated students.

Granted that there is some overlap or blurring of the distinction between coaching studies in Table 1 and instructional studies in Table 2, the two types seemed sufficiently different to warrant separate treatment. Accordingly, overall averages were not computed for the total combined set of studies because possible differences in impact might thereby be obscured. In contrast, Slack and Porter (1980) have chosen to combine both types of studies in a single table and to report overall weighted average score increases of coached groups over uncoached control groups or, when control groups were not utilized, over norm comparison groups. This assumes that the adjustments applied to the score gains in studies lacking control groups were large enough to correct appropriately for the experiential growth of self-selected students that would have occurred regardless of the program. And indeed, the adjustments made by Slack and Porter, being derived from normative data of dubious relevance to the preparatory school students in question, were smaller than those based on the average of four adjustments as applied in our Table 2. Because of this and various oversights in their table, the combined weighted averages of 29 points for SAT-V and 33 points for SAT-M that they report are more like 22 and 28 points, respectively (Messick, 1980, pp. 24–26). Such combined averaging is misleading not only because the adjustments may be questionable, but because the combined averages obscure important differences between the special preparation programs in the two types of studies. As indicated earlier, a comparison of the brief program descriptions in Tables 1 and 2 reveals that the programs in studies lacking control groups happened also to be quite long-term and intensive with respect to student contact time, whereas the programs in control-group studies were relatively short-term and nonintensive. The former programs also entailed organized curriculum content and skill development as well as test review, whereas the latter programs tended to emphasize test review and practice exercises.

Rather than averaging across these two

Table 1
Average Difference Between Experimental and Control Groups in Studies of SAT Coaching Interventions

	Sample characteristics				SAT-Verbal			SAT-Math		
Study/design	School	Level	Sex	Characteristics of the special preparation	Difference[a]	Significance level[b]	Experimental/control *N*	Difference[a]	Significance level[b]	Experimental/control *N*
Dyer (1953)/control, different school	Private	High school seniors	M	12 30-60-min sessions for Verbal, 5 60-90-min sessions for Math	4.6	<.05	225/193	12.9	<.01	225/193
French (1955)/control, different school	Public	High school seniors	M and F	10 Verbal and 10 Math coaching sessions using ETS item materials	18.3	<.01	161/158	6.2	<.01	161/158
French (1955)/control, different school	Public	High school seniors	M and F	4½ hours of vocabulary coaching, 10 sessions of Math coaching using ETS item materials	5.0	<.05	110/158	18.0	<.01	161/110
Dear (1958)/control, same and different schools	Public and private	High school seniors	M and F	Approximately 6 weekly 2-hour, 2-person coaching sessions plus 1 hour of homework each week	−2.5	*ns*	60/526	21.5	<.01	60/526
Dear (1958)/control, same and different schools	Public and private	High school seniors	M and F	Approximately 12 weekly 2-hour, 2-person coaching sessions plus 1 hour of homework each week	[c]	[c]		23.6	<.01	71/116
Frankel (1960)/control, same school statistically matched	Public	High school seniors	M and F	10 3-hour, 25-person coaching sessions	8.4	*ns*	45/45	9.4	*ns*	45/45
Whitla (1962)/control, statistically matched	Public and private	High school seniors	M and F	Proprietary coaching school, 5 2-hour sessions plus homework in Verbal and Math	11.0	*ns*	52/52	−5.3	*ns*	50/50
Roberts & Oppenheim (1966)/randomized	Public	High school juniors	M and F	7½ hours of programmed instruction in test taking and in Verbal and Math content	14.4[d]	<.05	154/111	8.1[d]	*ns*	188/122
Evans & Pike (1973)/randomized	Public	High school juniors	M and F	Test-taking skills and math content, 7 [illegible]-hour sessions, 21 hours of homework	No coaching for SAT-Verbal			16.5	<.05	288/129
Alderman & Powers (1980)/randomized	Public and private	High school juniors	M and F	Varied strategies, at eight schools, emphasizing reading and analogies, 5–45 hours	8.4	<.05	239/320	No coaching for SAT-Math		

Table 1 (*continued*)

Study/design	Sample characteristics			Characteristics of the special preparation	SAT-Verbal			SAT-Math		
	School	Level	Sex		Difference[a]	Significance level[b]	Experimental/control *N*	Difference[a]	Significance level[b]	Experimental/control *N*
FTC (1979, see also Stroud, 1980)/control, test-score files	Public and private	High school juniors and seniors	M and F	School A: 40 hours commercial coaching	31.7[e]	<.01	393/1729	24.9[e]	<.01	393/1729
				School B: 24 hours commercial coaching	5.2[e]	*ns*	163/1729	7.5[e]	*ns*	163/1729
Average weighted by size of experimental sample					14.3[f]			15.1[f]		

Note. SAT = Scholastic Aptitude Test; M = male; F = female; ETS = Educational Testing Service; *ns* = nonsignificant.

[a] The coaching effects are intercept differences between regression lines for experimental and control groups or (for Frankel, 1960; Whitla, 1962; Roberts & Oppenheim, 1966; and Pike & Evans, 1972) average score increases of experimental over control groups, both weighted in the case of multiple experimental or control groups by their respective sample sizes.

[b] As shown for coaching effects reported in original text.

[c] Not calculated; variances and regression slopes differed significantly for experimental and control groups.

[d] This study employed the Preliminary Scholastic Aptitude Test as both pre- and posttest; the averages shown have been converted to the SAT score scale ranging from 200 to 800 points.

[e] Weighted mean score effects pooling juniors and seniors across test administration years from Stroud's reanalysis of Federal Trade Commission (FTC) data.

[f] If Rock's (1980) estimates for FTC School A of 16.9 for SAT-Verbal and 30.6 for SAT-Math ($N = 192$) are substituted for Stroud's (1980) estimates, the weighted averages become 9.7 for Verbal and 14.5 for Math.

types of studies, which inevitably precipitates arguments about the appropriate size of the score effects to be included from the uncontrolled studies, let us instead *rank* all the studies in order of the reported treatment versus control group contrasts and when control groups are not available in order of the reported score increases. That is, for SAT-V, Pallone's (1961) long-term and summer programs would be ranked 1 and 2, respectively, followed by Marron's (1965) four groups, and so forth. This procedure grants that the score effects in the studies lacking control groups are larger in an ordinal sense than those in the control-group studies, but it takes no position with respect to how much larger. If the programs are then also ranked in terms of the number of student contact hours involved and a Spearman rank-order correlation coefficient computed, the rank correlation is found to be .62 across 24 studies for SAT-V and .74 across 15 studies for SAT-M. Both coefficients are significant at the .01 level. If 5 particularly suspect studies showing signs of probable biases in the score effects estimated are deleted from the calculations for SAT-V, the new correlation is .77 across 19 studies, which is also significant at the .01 level. In this latter calculation, Schools A, B, and D from Alderman and Powers (1980) were eliminated because of control-group score decreases, as was the Roberts and Oppenheim (1966) study for the same reason, and the Coffman and Parry (1967) study was dropped because of treatment-group score decreases and other indications of the low relevance to the SAT of both the accelerated reading program studied and the samples of enrolled college students employed. Further, if the Rock (1980) estimate taking account of differential group growth rates is substituted for the Stroud (1980) estimate of SAT-V effects at School A in the FTC [illegible]) study, the resulting rank correlations are .60 across 24 studies and .76 across 19 studies. If one suspect study, namely, Roberts and Oppenheim, is deleted from the calculations for SAT-M, the new correlation is .71, which is also significant at the .01 level. These rank-order correlations are summarized in Table 3. It should be noted that although the various coaching programs required dif-

Table 2
Adjusted Average Score Gains in Studies of SAT Instructional Interventions Without Control Groups

Study	Sample characteristics			Characteristics of the special preparation	SAT-Verbal		SAT-Math	
	School	Level	Sex		Adjusted average score increase[a]	*N*	Adjusted average score increase[a]	*N*
Pallone (1961)	Private	High school seniors and graduates	M	90-min. daily instruction and practice in developmental reading skills over 6 weeks	81	20+		
Pallone (1961)	Private	High school seniors and graduates	M	50-min. daily instruction and practice in developmental reading skills, with stress on logical inference and analogic analysis over 6 months	68 (43)[b]	80−		
Marron (1965)	Private	High school seniors and graduates	M	Full-time daily sessions aimed at verbal and mathematical content and test-taking skills over 6 months	Group 1 54 Group 2 33 } 35 Group 3 24 Group 4 12	83 600 5 26	Group 1′ 59 Group 2′ 53 Group 3′ 46	232 405 78
Coffman & Parry (1967)[c]	Public	College freshmen	M and F	6-hours weekly of instruction in accelerated reading over 8 weeks	4[d]	19		
Average weighted by size of experimental sample					38 (36)[b]		54	

Note. SAT = Scholastic Aptitude Test; M = male; F = female.

[a] To estimate instructional or program effects, average score increases in the Pallone (1961) and Marron (1965) studies were adjusted by the average of four adjustments, those suggested by (a) the authors of the original articles—Pallone suggested 35 points on SAT-Verbal as normal expectation of gains during the final year of secondary school (15 points for the 5-month interval between tests in the short-term program), and Marron suggested 24 and 26 points, respectively, for SAT-Verbal and SAT-Math as typical gains for high school seniors over 6 months; (b) Slack and Porter (1980)—average gains in national administrations of junior-to-senior year retesters having the same initial average score levels as Pallone's and Marron's groups; (c) Pike (1978)—average gains of control students in superior schools from other studies of proprietary programs; as well as, (d) average gains of control students in other studies who have average initial score levels roughly comparable to those of Pallone's and Marron's groups.

[b] Due to discrepancies in Pallone's tables, there is some uncertainty as to whether the average adjusted score effect in his long-term program should be 68 points or 43 points (see page 195).

[c] The 15-week program in Coffman and Parry was not included because the 29-point mean decrease in scores was considered atypical and possibly indicative of motivational and test-administration problems.

[d] The two 8-week programs in Coffman and Parry were combined, but adjustments were made only by the Slack and Porter procedure, which attenuated by only a few points an already tenuous effect. None of the suggested comparison groups of SAT takers appeared to provide even remotely reasonable yardsticks for gauging score gains of students already enrolled in a college not requiring the SAT.

ferent and usually unknown amounts of homework, this rank-correlation procedure tacitly assumes that the amount of homework in each case was roughly proportional to the number of student contact hours, so that the overall orderings would not be markedly changed if homework were taken into account.

In interpreting these sizable monotonic relationships between student contact time and score effects, it must be remembered that these are rank-order correlations between average values of different samples or groups, and correlations between averages are typically much higher than correlations between individual differences within groups. Furthermore, these rank correlations are dominated by the relative consistencies between the two types of studies with respect to the ranking variables—that is, the control-group studies are all relatively low in both student contact time and score effects, whereas the uncontrolled studies are all relatively high in both student contact time and score effects. It must also be remembered that the relatively high-contact programs entailed structured curricula emphasizing content knowledge and skill development, whereas the relatively low-contact programs emphasized test review and practice. With this confounding of program characteristics in mind, it appears that increases in student contact time (possibly serving as a proxy for increasing curriculum emphases on content knowledge and skill development) are systematically associated on the average with increases in SAT scores. However, even though the time dimension is covered in only a fragmentary fashion by the available studies, when the magnitude of (adjusted) score effect is plotted against student contact time, the relationship does not appear to be linear. Since the form of this relationship might enable us to predict the rough magnitude of score effects expected to be associated with coaching programs entailing given levels of student contact time, these data plots warrant careful examination.

Diminishing Returns in Coaching Effects

Attempting to approximate the general form of the function relating score effects to student contact time in coaching programs is difficult with the available data for two reasons: One is the fragmentary way in which the time dimension is covered by the existing studies, especially in the range between 30 hours and 300 hours, the latter representing half-time devoted to either Verbal or Math coaching in a 6-month full-time preparatory school program; the other is the inevitable variability or noise present in studies having methodological flaws. Barring the collection of new data, little can be done about the first difficulty. But with respect to the second, we can attempt to mute the effect of noise to some extent by excluding from consideration, at least initially, data from studies already identified as being particularly suspect—namely, data from Roberts and Oppenheim (1966), from Coffman and Parry (1967), and from Schools A, B, and D of Alderman and Powers (1980). Although these suspect data points will not be used to fit any functions, they will be plotted subsequently to see how aberrant they are with respect to functions based on the remaining data.

For somewhat different reasons, the data for Pallone's (1961) summer and long-term programs will also be excluded initially and subsequently plotted in relation to identified functions in the same way. Here the primary reason is uncertainty about how to estimate student contact time for these programs. Pallone's long-term program is described as 50 minutes of daily instruction over 6 months or approximately 100 hours, but since the students were also attending preparatory school full-time during this period, it seems more appropriate in the absence of control groups to estimate their contact time for Verbal preparation at 300 hours, as was done for Marron's (1965) preparatory school students. There is also some uncertainty, arising from discrepancies in Pallone's own tables, as to whether the average adjusted score effect for his long-term program is 68 points or 43 points (see Table 2). Moreover, the summer program is described as 90 minutes of daily instruction over 6 weeks, or approximately 45 hours, but since the students might also have been attending other summer preparatory courses during this period, it is impossible in the absence of information

Table 3
Correlations Between Rank Order of Score Effect and Rank Order of Student Contact Time for Studies of SAT Interventions With and Without Control Groups

Verbal[a]			
Study	Student contact time (hours)	Rank time	Rank score effect
Dyer (1953)	10[b]	13	15
French (1955)			
Verbal and Math	8.3[c]	14	8
Vocabulary	4.5	19	14
Dear (1958)	6	15.5	18
Frankel (1960)	15	9	12
Whitla (1962)	5	17.5	11
Alderman & Powers (1980)[d]			
School C	10.5	12	19
School E	6	15.5	17
School F	5	17.5	16
School G	11	11	9
School H	45	6.5	10
Pallone (1961)			
Short	45	6.5	2
Long	100	5	1
Marron (1963)			
Group 1	300	2.5	3
Group 2	300	2.5	4
Group 3	300	2.5	5
Group 4	300	2.5	6
FTC (1979; Stroud, 1980)			
School A	20	8	7[e]
School B	12	10	13[e]
Rank-order correlation		.77 (19 studies)[f]	
Roberts & Oppenheim (1966)[g]	3.8		
Alderman & Powers (1980)			
School A	7		
School B	10		
School D	10		
Coffman & Parry (1967)[h]	48		
Rank-order correlation		.64 (24 studies)[f]	

Math[a]			
Study	Student contact time (hours)	Rank time	Rank score effect
Dyer (1953)	8.3[i]	11.5	9
French (1955)[j]	8.3	11.5	10.5
Dear (1958)			
Long	12	9.5	6
Short	6	13	7
Frankel (1960)	15	8	12
Whitla (1962)	5	14	14
Evans & Pike (1973)			
Group QC	21	5	10.5
Group DS	21	5	8
Group RM	21	5	4
Marron (1965)			
Group 1[k]	300	2	1
Group 2[k]	300	2	2
Group 3[k]	300	2	3
FTC (1979; Stroud, 1980)			
School A	20	7	5[e]
School B	12	9.5	13[e]
Rank-order correlation		.71 (14 studies)	
Roberts & Oppenheim (1966)[g]	3.8		
Rank-order correlation		.74 (15 studies)	

Table 3 (*continued*)

Note. SAT = Scholastic Aptitude Test. QC = Quantitative Comparisons. DS = Data Sufficiency. RM = Regular Math.

[a] When only a total student contact time was available, it was assumed that half the time was devoted to Verbal and half to Math.

[b] Each of 12 exercises was estimated to require a 50-min. class period.

[c] Each of 10 exercises was estimated to require a 50-min. class period.

[d] Schools A, B, and D were suspect because of control-group score decreases and hence were omitted in these calculations based on 19 studies; they are included in the later calculations based on all 24 studies.

[e] Ranks based on weighted mean score effects pooling juniors and seniors across test administration years from Stroud's reanalysis of Federal Trade Commission (FTC) data.

[f] If the Rock (1980) estimate taking account of differential group growth rate is substituted for the Stroud (1980) estimate of SAT-Verbal effects at School A in the FTC study, the resulting correlations are .76 across 19 studies and .60 across 24 studies.

[g] Roberts and Oppenheim was suspect because of control-group decreases and hence was omitted from the calculations for Verbal based on 19 studies and for Math based on 14 studies.

[h] Coffman and Parry was suspect because of treatment-group score decreases and other indications of motivational problems and hence was omitted from the calculations based on 19 studies. The 25-week program was dropped altogether because of the treatment-group score decreases, and the two 8-week programs were combined, yielding a weighted average score effect of 6.5.

[i] Each of five double-length exercises was estimated to require two 50-min. class periods.

[j] French's Math coaching group was contrasted with two control groups, one in a school having no coaching and one in a school having vocabulary coaching; these two comparisons were combined, yielding a weighted average score effect of 11.

about their programs to estimate what the effective student contact time for Verbal preparation might have been. If a control group had been available of similar students attending the same summer school but who were not taking Pallone's course, then the average net score increase for Pallone's students over the control students could have been associated with the 45 hours of student contact time in Pallone's course description. In addition, since Pallone's summer course was a *pilot* program involving only about 20 students, it seemed prudent to give it little or no weight in fitting functions.

With these exclusions, 17 data points remain for SAT-V score effects and 14 data points for SAT-M effects (see Table 3). When these (adjusted) values are plotted separately for Verbal and Math scores as a function of their associated student contact times, the resulting arrays appear nonlinear. This may be demonstrated by contrasting the linear function that best fits all of the Verbal (or Math) data points with the linear function that best fits only those data points toward the low end of the time dimension, that is, all but the full-time programs. If the overall relationship were approximately linear, these two regression lines would be quite similar to each other. But their slopes, as we shall see later, indeed differ by a factor of 4 or 5, which makes an enormous difference over a time range of 300 hours.

Given this broad time range and the fact that the data points are clustered at the two extremes, the nature of the nonlinear relationship is better revealed analytically than visually. The linear least squares regression equation for the 17 Verbal studies is $V = 7.336 + .079T$, where V is SAT-V score effect and T is time in hours ($N = 17$, $r = .66$). This equation is dominated by Marron's (1965) four Verbal groups at 300 hours of student contact time, for which the predicted score effect of 31 points is close to their adjusted weighted average (see Table 2). But the predicted values at the low end of the time scale are out of line with the data—e.g., the predicted score effect for 1 hour of contact time is 7.4 points; for 10 hours, 8.1 points; for 20 hours, 8.9 points; for 30 hours, 9.7 points; for 50 hours, 11.3 points. The linear regression equation for the 13 data points toward the low end of the time scale, ignoring Marron's four groups for the moment, is $V = 4.322 + .335T$, ($N = 13$, $r = .37$). The predicted values toward the low end of the scale are now more in line with the data (e.g., 4.7 score points for 1 hour of contact time, 7.7 for 10 hours, 11.0 for 20 hours, 14.4 for 30 hours, 21.1 for 50 hours), but the predicted values toward the high end are out of line (e.g., the predicted score effect for 300 hours of contact time is 105 points, which is three times as large as the adjusted weighted SAT-V effect for Marron's four groups and almost twice as large as their unadjusted weighted average). Thus, the regression line based on data from coaching programs entailing less than 50 hours of student contact time projects an expected score effect for programs of 300 contact hours that is two to three times higher than the average effect actually associated with those 300-hour programs in practice. Consequently, as student contact time increases, there appear to be diminishing returns in SAT-V score effects associated with coaching.

A similar picture emerges for SAT-M. The linear regression equation for the 14 Math studies is $M = 12.699 + .134T$, ($N = 14$, $r = .89$). Again, this equation is dominated by Marron's (1965) three Math groups at 300 hours of student contact time, for which the predicted score effect of 53 points is close to their adjusted weight average (see Table 2). But once more the predicted values at the low end of the time scale are out of line with the data—for example, 12.8 score points for 1 hour, 14.0 for 10 hours, 15.4 for 20 hours, 16.7 for 30 hours, and 19.4 for 50 hours. The linear regression equation for the 11 data points toward the low end of the time scale, ignoring Marron's three groups momentarily, is $M = 4.900 + .713T$, ($N = 11$, $r = .49$). The predicted values toward the low end of the scale, as they were for Verbal, are now more in line with the data (e.g., 5.6 score points for 1 hour, 12.0 for 10 hours, 19.2 for 20 hours, 26.3 for 30 hours, and 40.5 for 50 hours), but those toward the high end of the scale are out of line (e.g., the predicted score effect for 300 hours of contact time is 219 points, which is over four times as large as the adjusted weighted SAT-M effect for Marron's three Math groups and nearly three times as large as their unadjusted weighted average). Clearly, as student contact time increases, there appear to be diminishing returns in SAT-M score effects as well.

As is frequently the case in situations with diminishing returns, a logarithmic transformation of the time dimension provides a much better representation of the functional relationships, as is seen in Figure 1 for SAT-V and Figure 3 for SAT-M. In this formulation, all logs are to the base 10. The solid line in Figure 1 is the regression of SAT-V score effect on student contact time in log hours, $L(T)$, based on all 17 data points discussed previously; its equation is $V = -6.587 + 15.155L(T)$, ($N = 17$, $r = .70$). The dashed regression line in Figure 1 is based on the 13 data points toward the low end of the time scale, omitting the values for Marron's (1965) four Verbal groups; its equation is $V = -7.768 + 16.418L(T)$, ($N = 13$, $r = .47$). In contrast to the discrepant regression lines obtained when a linear fit in real time was attempted for 17 and 13 data points, these two regression lines in log time are very similar to each other—for example, the predicted SAT-V score effect corresponding to 300 hours of student contact time (2.477 log hours) is 31 points by the first equation and 33 points by the second. It should be noted that these predicted values, even though based on an unweighted least squares solution, closely correspond to the adjusted score effect for Marron's largest group ($N = 600$) and that his other three more deviant groups entailed much smaller sample sizes ($N = 83$, 5, and 26 in descending order of score effect). According to these logarithmic functions, an estimated zero SAT-V score effect is obtained at somewhat less than 3 hours, suggesting that additional data points in this region may require more curvature in the function or, more likely, that a threshold or critical mass of coaching effort must be attained before positive score effects are exhibited.

The triangle in Figure 1 corresponds to Rock's (1980) estimate of the SAT-V score effect taking differential group growth rates into account for students at School A in the FTC (1979) study. Even though these two logarithmic regression equations were based

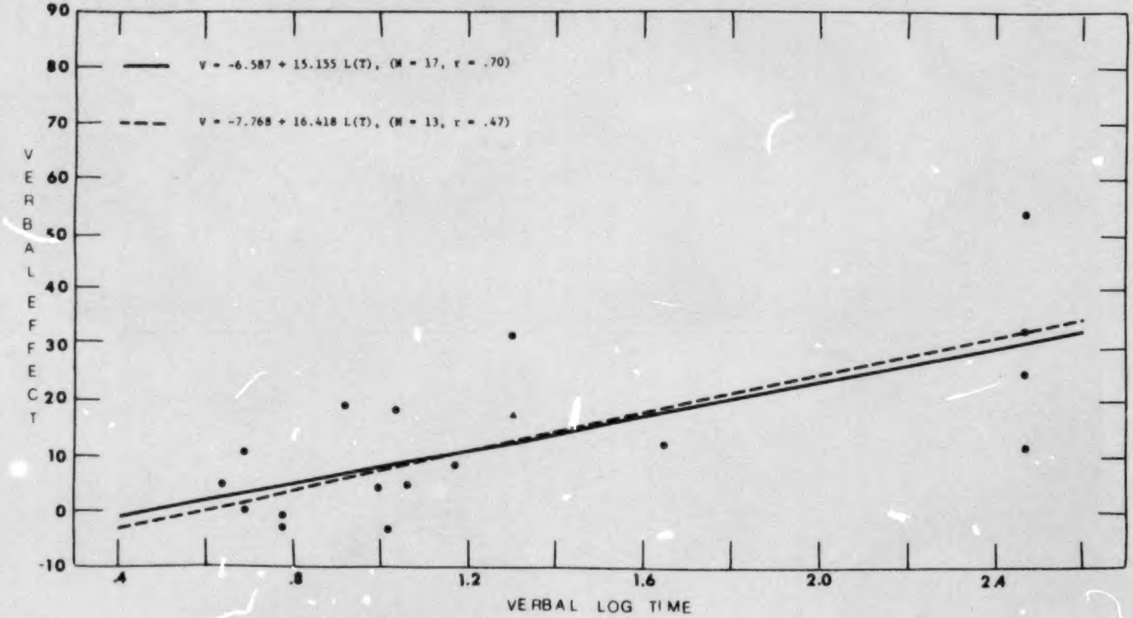


Figure 1. Regressions of SAT-Verbal score effects on student contact time in logs, with those studies identified as suspect omitted.

on Stroud's (1980) estimate for School A, the Rock estimate is clearly more consistent with both regression lines. The regression equation substituting Rock's estimate of the SAT-V effect for School A in the FTC data for Stroud's estimate is $V = -7.544 + 15.221\,L(T)$, ($N = 17$, $r = .74$), which is very similar to the regression equation based on Stroud's estimate but, by virtue of the greater congruence of Rock's value, reflects a somewhat higher correlation among the data points.

Figure 2 expands the data base represented in Figure 1 by adding the data points excluded initially. The five circles in Figure 2 correspond to the five suspect data points of Roberts and Oppenheim (1966), Coffman and Parry (1967), and Schools A, B, and D from Alderman and Powers (1980). It can be seen that these points are somewhat noisier than the rest, the correlation between SAT-V effect and log time being .59 for 22 data points but .70 for 17 points. However, as is apparent in Figure 2, the regression equation based on 22 data points does not differ dramatically from the other equations: $V = -2.405 + 12.566L(T)$, ($N = 22$, $r = .59$).

Finally, the squares in Figure 2 correspond to the two discrepant estimates of adjusted SAT-V effects for Pallone's (1961) long-term program (see Table 2). When plotted at 300 hours (2.477 log hours), as previously argued, the smaller of these estimates (43) is quite consistent with the logarithmic regression system for 13, 17, and 22 data points, whereas the larger estimate (68) is deviant but not incompatible. The logarithmic regression equation for the 23 data points including the 43-point estimate at 300 hours of student contact time is $V = -4.025 + 14.253L(T)$, ($N = 23$, $r = .66$); substituting the larger 68-point estimate at 300 hours for the smaller value, the equation becomes $V = -6.673 + 17.089L(T)$, ($N = 23$, $r = .66$). The first equation is quite similar to the three plotted in Figure 2 and the second equation is not markedly different. As compared with the Figure 2 estimates of 33, 31, and 29 points for a benchmark 300 hours of student contact time, the regression equations including Pallone's data yield 31 and 36 points. Even when Pallone's estimated score effects are plotted at 100 hours of student contact time (tacitly treating Pallone's course in isolation from other verbal preparation likely accruing from full-time school attendance), the respective regression lines are still quite comparable to those shown in Figure 2: For 43 score-effect points, $V = -3.618 + 14.166L(T)$, ($N = 23$, $r = .62$), and for 68 score-effect points, $V = -4.973 + 16.051L(T)$, ($N = 23$, $r = .59$). These equations yield estimates of 31 and 35 points, respectively, for the benchmark of 300 hours of student contact time. Nevertheless, if plotted at 100 hours, these data points for Pallone's long-term program are quite deviant, particularly the larger value of 68 points, but the reported score effect for Pallone's summer pilot program appears to be aberrant in this logarithmic formulation for any plausible range of adjustments for experiential growth or any plausible range of student contact time. In sum, the logarithmic regression system relating SAT-V score effect to log contact time based on 17 data points appears to be quite robust under the addition of the noisier values available from a half-dozen other studies all of which are questionable for one reason or another.

The solid line in Figure 3 is the regression of SAT-M score effect on student contact time in log hours based on 14 data points (see Table 3); its equation is $M = -14.072 + 26.646L(T)$, ($N = 14$, $r = .91$). The dashed regression line in Figure 3 is based on the 11 data points toward the low end of the time scale, omitting the values for Marron's (1965) three Math groups; its equation is $M = -7.911 + 20.775L(T)$, ($N = 11$, $r = .51$). In contrast to the marked discrepancy in regression lines obtained when a linear fit in real time was attempted for 14 and 11 points, these two regression lines in log time are very similar to each other although they are not quite as consonant as the corresponding lines in the Verbal data, as might be expected because of the smaller samples of data points. For example, the predicted SAT-M effect corresponding to 300 hours of student contact time (2.477 log hours) is 52 points by the first equation and 44 points by the second, an 8-point difference, which is a far cry from the 166-point

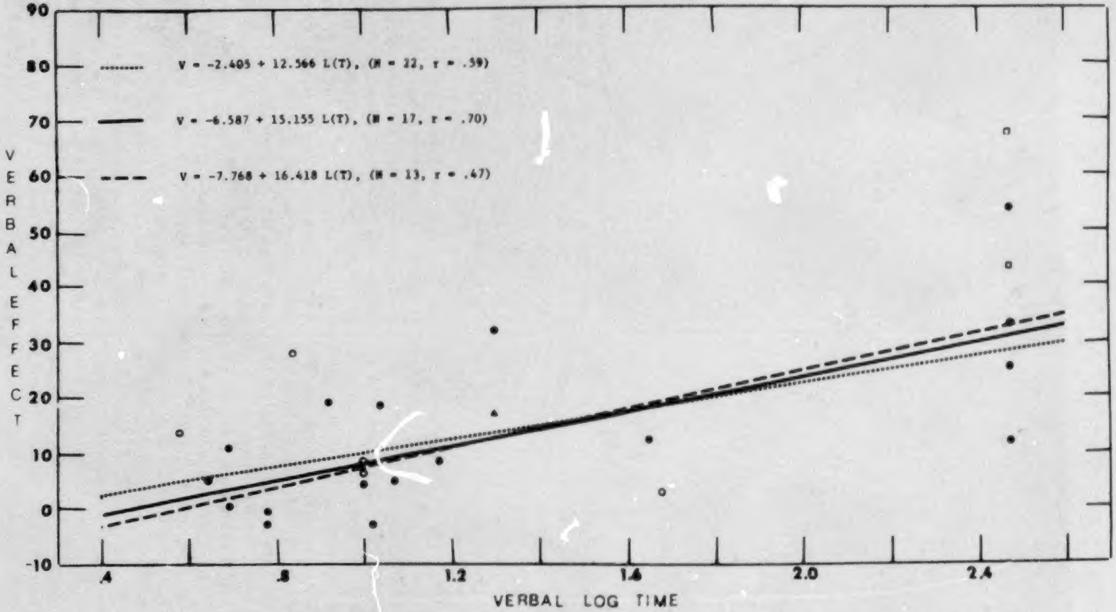


Figure 2. Regressions of SAT-Verbal score effects on student contact time in logs for all Verbal coaching studies.

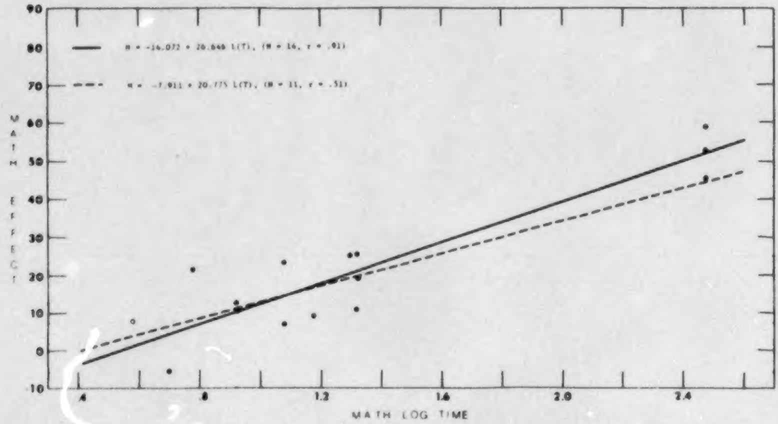


Figure 3. Regressions of SAT-Math score effects on student contact time in logs for all Math coaching studies.

discrepancy in the same 300-hour estimates based on a linear fit in real time to 14 and 11 data points. The one circle in Figure 3 represents the Roberts and Oppenheim (1966) data, which can be seen to be quite consistent with these logarithmic regressions; the regression equation including this additional data point is $M = -12.451 + 25.763L(T)$, ($N = 15$, $r = .91$). Again, according to these logarithmic functions, an estimated zero SAT-M score effect is obtained at around 3 hours or somewhat less, indicating that the threshold of coaching effort needed for positive score effects, if such be its import, is roughly comparable for Verbal and Math—although the slope of the Math function is more than half again as steep as that of the Verbal function.

In regard to appraising the goodness of fit of the data points to the proposed logarithmic model, it can be seen in Figures 1 and 2 that the largest residuals for SAT-V are associated with Marron's (1965) Groups 1 and 4, with Stroud's (1980) estimate for FTC (1979) School A, and with three of the five suspect studies. This is attested to by noting that the root-mean-square residual for 22 data points is 10.9, which is reduced to 10.5 by eliminating the five suspect studies, further reduced to 8.4 by dropping Marron's four groups, and still further reduced to 6.6 by substituting Rock's (1980) FTC School A estimate for Stroud's. The goodness of fit for SAT-M as revealed in Figure 3 is even better. The root-mean-square residual for 14 data points is 7.3 whereas that for 11 data points is 7.6. Although it would be desirable to evaluate the size of residuals in relation to the standard error of their associated score-effect estimate, many of the studies did not report standard errors, nor could they be readily calculated from the information given. Nonetheless, there was some indication that deviations from the model were at least partially a function of sampling error because the larger residuals tended to be associated with studies having smaller sample sizes. The rank-order correlation between the absolute value of SAT-V residuals and the size of their corresponding treatment-group sample was −.30 for the regression model based on 17 studies (still using Rock's estimate), whereas the rank-order correlation for SAT-M residuals based on 14 studies was −.56. Although the former coefficient is not statistically significant, the latter value is. Overall, then, the proposed logarithmic model, however adventitious it may appear at first glance, offers a plausible hypothesis for the relationship between increasing student contact time and increasing score effects in programs of coaching and instruction for the SAT.

To recapitulate briefly, substantial rank-order correlations were noted between SAT score effects in coaching studies and the level of student contact time entailed in each coaching program (see Table 3). This regularity of monotonic relationship appears not to be linear, however, but approximately logarithmic, at least within the limitations of the fragmentary data. If this suggested logarithmic relationship has substance, then each additional increase in SAT scores associated with coaching may require geometrically increasing amounts of student contact time and of all the curricular effort that contact time may be proxy for. For example, according to the logarithmic function for SAT-V based on 17 data points, an average score increase of 10 points on a 200- to 800-point scale would be expected to be associated with approximately 12 hours of student contact time in a Verbal coaching program, but 20 score points would require 57 hours, 30 score points would require 260 hours, and for 40 SAT-V score points a total of 1,185 hours would be required (although the latter extrapolation is far beyond the range of the data). Similarly, for the SAT-M function based on 14 data points: An average score increase of 10 points would be expected to be associated with approximately 8 hours of student contact time, but 20 score points would require 19 hours, 30 score points would require 45 hours, and to attain 40 SAT-M score points would require a total of 107 hours. The fact that Math entails less student contact time than Verbal for a given amount of score increase is consistent once again with the expectation that Math, being more curriculum related than Verbal, should be relatively more responsive to instructional intervention.

It must be emphasized that these functions apply to existing data and that the available studies all involved students who to a large degree were motivated to increase their test scores through coaching or special preparation. Such motivated students are likely to have been highly task oriented, and under such circumstances it is not unreasonable that student contact time should be directly related to average test score increases. We have also cautioned that increases in student contact time are confounded in these studies with increasing curriculum emphases on content knowledge and skill development, so that other program characteristics that contact time may be proxy for should also be taken into account in interpreting the relationship and in shaping expectations. Indeed, the terms *time* and *method* were paired in the title of this article partly to underscore that student contact time and aspects of coaching method are confounded in the available coaching studies and partly to highlight the importance of recurrent flaws in experimental method that serve to obscure the functional role of contact time in understanding coaching effects. Speaking of methodological flaws, we must finally also consider the possible role of selection bias in interpreting the basis of the observed relationship, because many of the longer-term programs such as the preparatory schools in Marron's (1965) study and the commercial coaching schools in the FTC (1979) study were not only associated with larger score effects but were also highly subject to self-selection bias.

With these caveats in mind, it appears likely that improvement of the comprehension and reasoning skills measured by the SAT, when it occurs, is a function of the time and effort expended and that each additional score increase may require increasing amounts of time and effort, probably geometrically increasing amounts. If this is the case, the time required to achieve average score increases much greater than 20 to 30 points (on a 200- to 800-point scale) rapidly approaches that of full-time schooling, especially in Verbal; hence it quickly becomes unfeasible to augment the already full-time requirements of secondary school with sufficient additional contact time devoted to coaching to obtain large improvements in comprehension and reasoning skills over those afforded by a high school education. As a consequence, the soundest long-range mode of preparation for the SAT would appear to be a secondary school program that integrates the development of thought with the development of knowledge.

References

Alderman, D. L., & Powers, D. E. The effects of special preparation on SAT-verbal scores. *American Educational Research Journal*, 1980, *17*, 239–253. (Also, CB RDR 78-79, No. 4, and ETS RR 79-1, Princeton, N.J.: Educational Testing Service, 1979.)

Anastasi, A. Abilities and the measurement of achievement. In W. B. Schrader (Ed.), *New directions for testing and measurement—Measuring achievement: Progress over a decade*. San Francisco: Jossey-Bass, 1980.

Bryk, A. S., & Weisberg, H. I. Use of the nonequivalent control group design when subjects are growing. *Psychological Bulletin*, 1977, *84*, 950–962.

Cochran, W. G. The use of covariance in observational studies. *Applied Statistics*, 1968, *17*, 270–275.

Coffman, W. E., & Parry, M. E. Effects of an accelerated reading course on SAT-V scores. *Personnel and Guidance Journal*, 1967, *46*, 292–296.

Dear, R. E. *The effect of a program of intensive coaching on SAT scores* (ETS RB 58-5). Princeton, N.J.: Educational Testing Service, 1958. (Reported in French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. *Educational and Psychological Measurement*, 1959, *19*, 319–330.)

Dyer, H. S. Does coaching help? *College Board Review*, 1953, *19*, 331–335. (Reported in French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. *Educational and Psychological Measurement*, 1959, *19*, 319–330.)

Evans, F. R., & Pike, L. W. The effects of instruction for three mathematics item formats. *Journal of Educational Measurement*, 1973, *10*, 257–272.

Federal Trade Commission, Boston Regional Office. *Staff memorandum of the Boston Regional Office of the Federal Trade Commission: The effects of coaching on standardized admission examinations*. Boston, Mass.: Author, September 1978.

Federal Trade Commission, Bureau of Consumer Protection. *Effects of coaching on standardized admission examinations: Revised statistical analyses of data gathered by Boston Regional Office of the Federal Trade Commission*. Washington, D.C.: Author, March 1979.

Frankel, E. Effects of growth, practice, and coaching on Scholastic Aptitude Test scores. *Personnel and Guidance Journal*, 1960, *38*, 713–719.

French, J. W. *The coachability of the SAT in public schools* (ETS RB 55-26). Princeton, N.J.: Educational Testing Service, 1955. (Reported in French, J. W., & Dear, R. E. Effect of coaching on an aptitude test. *Educational and Psychological Measurement*, 1959, *19*, 319–330.)

Marron, J. E. *Preparatory school test preparation: Special test preparation, its effect on College Board scores and the relationship of affected scores to subsequent college performance*. West Point, N.Y.: Research Division, Office of the Director of Admissions and Registrar, United States Military Academy, 1965.

Messick, S. *The effectiveness of coaching for the SAT: Review and reanalysis of research from the fifties to the FTC* (ETS RR 80-8). Princeton, N.J.: Educational Testing Service, 1980.

National Education Association. *Measurement and testing: An NEA perspective*. Washington, D.C.: Author, 1980.

Pallone, N. J. Effects of short-term and long-term developmental reading courses upon S.A.T. verbal scores. *Personnel and Guidance Journal*, 1961, *39*, 654–657.

Pike, L. W. *Short-term instruction, testwiseness, and the Scholastic Aptitude Test: A literature review with research recommendations* (CB RDR 77-78, No. 2, and ETS RB 78-2). Princeton, N.J.: Educational Testing Service, 1978.

Pike, L. W., & Evans, F. R. *The effects of special instruction for three kinds of mathematics aptitude items* (CB RDR 71-72, No. 7, and ETS RB 72-19). Princeton, N.J.: Educational Testing Service, 1972.

Roberts, S. O., & Oppenheim, D. B. *The effect of special instruction upon test performance of high school students in Tennessee* (CB RDR 66-7, No. 1, and ETS RB 66-36). Princeton, N.J.: Educational Testing Service, 1966.

Rock, D. A. *Disentangling coaching effects and differential growth in the FTC commercial coaching study* (ETS RR 80-11). Princeton, N.J.: Educational Testing Service, 1980.

Slack, W. V., & Porter, D. The Scholastic Aptitude Test: A critical appraisal. *Harvard Educational Review*, 1980, *50*, 154–175.

Snow, R. E. Aptitude and achievement. In W. B. Schrader (Ed.), *New directions for testing and measurement—Measuring achievement: Progress over a decade*. San Francisco: Jossey-Bass, 1980.

Stroud, T. W. F. *Reanalysis of the Federal Trade Commission study of commercial coaching for the SAT* (ETS RR 80-10). Princeton, N.J.: Educational Testing Service, 1980.

Whitla, D. K. Effect of tutoring on Scholastic Aptitude Test scores. *Personnel and Guidance Journal*, 1962, *41*, 32–37.

Received October 8, 1980 ■

American Dental Hygienists' Association

Kathleen O. Smith, R.D.H., *President*
Marjorie J. Sharpe, *Executive Director*
Ben F. Miller III, *Deputy Manager, Federal Affairs*

November 25, 1981

Honorable Carl D. Perkins
Chairman
Subcommittee on Elementary,
Secondary and Vocational Education
2365 Rayburn Building
Washington, D.C. 20515

Dear Mr. Perkins:

At this time, the Association is not in favor of federal legislation to regulate the testing industry. There is still no evidence that such legislation is needed. The Association feels that the results of the National Academy of Sciences study on standardized testing, soon to be released, should be carefully assessed by the subcommittee prior to recommending enactment of federal legislation. The Association believes that much more evidence favoring the need for federal regulation of the testing industry should be compiled than the subcommittee has already included in the hearings record.

Recent testimony presented to the subcommittee reinforces the view that the threat of possible regulation has already impacted on the testing industry and currently, disclosure of test information to students and parents is already occurring on a voluntary basis. Educational Testing Service (the giant in the field) has reversed itself and is complying with provisions of the legislation on its own initiatives. The law school association has set a full-disclosure policy for the LSAT's, etc. In addition to the voluntary actions taken by some of the majors in the business, the climate in Congress and the Executive Branch of the federal government is clearly anti-regulatory. In short, it is our view that the real or remote possibility of federal regulation has already produced reforms in the testing industry of a substantive nature, to the point that enactment of a Truth-in-Testing law now seems redundant. The absence of scheduled FTC witnesses on the second day (November 5) of the recent hearings was probably not inadvertant.

Washington Office • Suite 1006 • 1101 17th Street N.W. • Washington, D.C. 20036 • (202) 833-3[illegible]24

The Dental Hygiene Aptitude Test is the only currently available instrument that is specifically designed for applicants seeking admission to dental hygiene schools. This test provides information to the applicant on his or her standing compared to others with comparable education. Evidence shows that this type of information is useful to admissions officers.

If this proposed legislation is enacted, the Dental Hygiene Aptitude Test would be adversely affected, since the cost of developing new examinations each year would be prohibitive and diminish its desirability to institutions and to prospective dental hygiene students. To preverse the integrity of the test, at least three test forms would have to be developed annually. Costs would triple in order to: maintain a suitable size item pool, cover costs for extra printing, ship additional test booklets and construct new subtests. These costs, of course, would have to be passed along to the consumer; i.e., high school graduates applying for admission to dental hygiene schools.

The DHAT scores are not the only criteria used by institutions in admissions screening. Personal interviews, grade point average, other test scores and life experiences are also part of the selection process. The DHAT can offer an applicant who has long been out of school information on the likelihood of academic success; it can also identify individuals who may require scholastic remediation. It is obvious that the DHAT can be a useful admissions tool.

The Association believes that any legislation at this point would be detrimental to the DHAT and to the colleges that use this instrument as an initial screening device. The gain in predicting academic success, as compared to selection of students at random, is large enough to warrant its use. At this point, the Association believes that the pro-regulation side of the debate has failed to make a convincing case for regulating an industry which maintains the highest standards of professional integrity in the public interest.

Sincerley,

Kathleen D. Smith
Kathleen D. Smith, RDH, MS
President, ADHA

cc: Subcommittee on Elementary, Secondary and Vocational Education
ADHA Board of Trustees

SOUTH MILWAUKEE PUBLIC SCHOOLS

SOUTH MILWAUKEE SENIOR HIGH SCHOOL
1801 Fifteenth Avenue — South Milwaukee, Wisconsin
Tel.: 768-6318 53172

R. W. Schaus, *Superintendent*
N. S. Klein, *Principal*
D. J. VanderVelden, *Asst. Prin.*
P. Strosyk, *Asst. Prin.*

December 3, 1981

Honorable Clement J. Zablocki
2183 Rayburn House Office Building
Washington, D.C. 20515

Dear Representative Zablocki:

It would be appreciated if the enclosed letter were forwarded to the committee that is considering "truth in testing" legislation.

Yours truly,

Joe Baer

Joe Baer
Director of Guidance

JB:do
Enclosure

SOUTH MILWAUKEE SENIOR HIGH SCHOOL
1001 Fifteenth Avenue — South Milwaukee, Wisconsin
Tel: 768-6315 53172

R. W. Schaus, *Superintendent*
N. S. Klein, *Principal*
D. J. VanderVelden, *Asst. Prin.*
P. Strozyk, *Asst. Prin.*

December 3, 1981

The College Board
888 Seventh Avenue
New York, New York 10019

Gentlemen:

The Guidance Department at South Milwaukee Senior High School is deeply disturbed by the trend to coach students for the S.A.T. When the College Board itself outlines a six weeks course for counselors to coach students, the situation is out of hand. No outside agency has the right to tell counselors how to use their time, and if our students are going to compete, it appears we must start coaching. Many schools, particularly large city schools, will not be able to provide counselor time to provide such courses. Again we "hurt" our minorities. Your latest statistics show that S.A.T. scores did not decline last year for the first time since 1968. How much of this statistical change is due to coaching?

Another disturbing trend in the College Board program is the tendency to test students at the tenth grade level for the PSAT/NMSQT and at both the eleventh and twelth grade for the S.A.T. The increasing costs and the pressure on students to take tests for practice are very negative factors. For years we believed College Board when they stated that coaching and practice did not significantly influence test scores. Now College Board advocates coaching and tests for practice.

We have always supported the College Board and its testing program in the past. This is no longer true and we now believe that if this trend toward coaching continues, perhaps we should eliminate testing as one of the criteria for college admissions since not all students will be coached or take practice tests.

Yours truly,

Joe Baer

Joe Baer
Director of Guidance

Eunice Hahn

Eunice Hahn
Counselor

Alan Hansis

Alan Hansis
Counselor

Wallace Swanson

Wallace Swanson
Counselor

JB:do
cc: Ann Fritts, President National Association of College Admissions Counselors
Kathy Bourne, President, Wisconsin Association of Secondary School and College Admissions Counselors
Senator William Proxmire
Senator Robert Kasten
Representative Clement Zablocki